(47)

Using Multiresolution Range-Profiled Real Imagery in a Statistical Object Recognition System

by

Asuman E. Koksal

B.S., Middle East Technical University (1996)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1998

Signature of Author .....
Department of Electrical Engineering and Computer Science
May 8, 1998

Certified by .....
Professor Jeffrey H. Shapiro
Professor and Associate Head of Electrical Engineering
Thesis Supervisor

Certified by .....
Dr. William M. Wells
Research Scientist, Artificial Intelligence Laboratory
Assistant Professor of Harvard Medical School
Thesis Supervisor

Accepted by .....
Professor Arthur C. Smith
Chairman, Department Committee on Graduate Students

Using Multiresolution Range-Profiled Real Imagery in a Statistical Object Recognition System

by

Asuman E. Koksal

Submitted to the Department of Electrical Engineering and Computer Science
on May 8, 1998, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Recognizing 3-D objects from range imagery has received considerable attention in the last few years. Laser radar range imagery is degraded by the combined effects of laser speckle and local oscillator shot noise, resulting in range anomalies and Gaussian noise in the local accuracy of the range measurements. Our objective was to develop a statistically optimum approach for doing model-based object recognition using low-resolution, noise-degraded laser radar range images. The object recognition system we developed consists of preprocessing, segmentation, feature extraction and alignment/scoring steps. For the preprocessor, we have employed the fast ML/EM algorithm, which is an essentially optimal anomaly suppression scheme. The resulting range profile is segmented using planar range profiling to estimate and isolate the target region from the background as accurately as possible. The feature extraction module provides relevant edge-based features that distil the essential characteristics needed to identify the target in the segmented range images. The alignment/scoring step estimates the pose of the target in the image, based on the posterior marginal pose estimation method (PMPE), and performs matching between image and model features. The output of the system is the value of the objective function of the PMPE matcher, which gives an indication of the degree of alignment between the image and each member of the object-model data base. The scores of alignment for each model are compared to find the type of the located target in the image. Performance results for the object recognition system are presented using laser radar data from the MIT Lincoln Laboratory Infrared Airborne Radar data release and 3-D CAD models that account for possible military targets that may be present on the site imaged by the laser radar. The performance of the system is also analyzed as a function of the sensor parameters to determine the effect of sensor physics capabilities on the recognition module.

Thesis Supervisor: Professor Jeffrey H. Shapiro

Title: Professor and Associate Head of Electrical Engineering

Thesis Supervisor: Dr. William M. Wells
Title: Research Scientist, Artificial Intelligence Laboratory
Assistant Professor of Harvard Medical School

Acknowledgments

I feel very fortunate to have Professor Jeffrey Shapiro as my thesis advisor. I am deeply thankful to him for his incredible guidance, support and understanding throughout this work. His clarity of thought, deep knowledge, enthusiasm for his work and his commitment to excellence was a great inspiration and will always set an example for me throughout my life.

I am also grateful to Dr. Sandy Wells for supervising my thesis. His contribution to this research is invaluable. I thank him so much for introducing me to his work on statistical object recognition, for offering his time for many useful discussions on this work and for helping me on many implementation details.

I would like to thank Professor Alan Willsky and everyone in the Stochastic Systems Group for letting me use their computer laboratory and for providing invaluable help and advice. I also thank my officemates Chen-Peang Yeang and Jeff Bounds for their friendship, help and stimulating conversations.

I greatly appreciate the support I have received from the U.S. government, through U.S. Army Research Office of Sponsored Research Grant DAAH04-95-1-0494.

The laser radar data used in this thesis was obtained from the MIT Lincoln Laboratory Infrarared Airborne Radar Program Data Release.

Finally, I am deeply grateful to my husband, Emre, for his unfailing love and patient support. His daily encouragement has given me the strength to overcome the challenges and difficulties. I would like to thank all my family members for their support, love and for always believing in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A coherent laser radar can produce range, intensity or Doppler images by raster scanning a field of view in 3-D pulsed imager mode. The goal of this research is to develop a target recognition system capable of detecting and recognizing military vehicles in range images provided by airborne laser radars. In particular, we will focus on using laser radar range imagery in a model-based, statistical object recognition system. Toward that end, it is effective to use an essentially optimal image processing mechanism that can act as the input stage for the recognition system to combat the degradation processes encountered in generating range imagery. The result is a laser-radar-based object recognizer.

Recognizing 3-D objects from range imagery has received considerable attention in the last few years. However, most approaches followed so far are applicable to high angular-resolution and high range-precision range images. In this thesis, we focus on developing an end-to-end object recognition system operating on noisy, low resolution range data to recognize an object and find its location throughout the image.

Research on the statistics of peak-detecting coherent laser radars has led to the development of techniques for performing target detection for 2-D imagers and 3-D imagers. Previous work by T.J. Green [1] has shown maximum-likelihood (ML) planar range profiling with the expectation-maximization (EM) algorithm to be a computationally simple and efficient procedure with good noise and anomaly suppression. I. Fung and D.R.

Greer [2]-[4] extended the planar range profiling work to fit a multiresolution basis at a sequence of increasingly fine resolutions to a laser radar range image. The EM algorithm was used to obtain the ML estimate for the range image. The multiresolution range profiling work was applied to 3-D real laser radar range imagery by means of a more powerful EM algorithm designed using the special structure of the Haar wavelet basis employed in the algorithm, which provides low computational complexity and excellent numerical robustness.

An effective feature-based object recognition system must be able to deal with variability in the positions of image features and the appearance of unexpected features due to background effects. This issue of uncertainty in feature location and detection suggests a statistical approach to object recognition. A statistical framework for object recognition was formulated by W.M. Wells [5], which provides explicit models for the uncertainties involved. By these probabilistic models, methods of statistical estimation were used in recognition of objects in high resolution video or synthetic range images.

The research in this thesis focuses on whether the preceding mechanisms can successfully be combined to process real, low resolution laser radar range images. The combined system will be applied to real laser radar range images in the IRAR Data Release from MIT Lincoln Laboratory. These images have low resolution compared to the previously used video images and synthetic range images. Moreover, there is an additional dimensionality reduction involved in multiresolution range profiling of data. The resulting resolution may not be sufficient to have adequate number of image features for the object recognition algorithm to work properly. The objective of this research is to find out if this approach works for real laser radar range data.

The object recognition system is extended to involve a data library of 3-D models, which will be used to identify the target in the real laser radar data among a collection of possible military targets. Moreover, as opposed to 2-D range truths used previously, use of 3-D models of specific military vehicles converts the object recognition module into a system that can recognize objects having 6 degrees of freedom of position. Our intent is to

obtain an object recognition algorithm that can be applied to real data with an optimal front end by which the effects of sensor physics capabilities on the feature extraction mechanisms and hence the object recognition performance can be well understood. In this way, it may be possible to backpropagate the feature accuracy requirements of the object recognition module to the performance requirements for the near-optimal front end processor. These requirements, in turn, would be backpropagated to determine the radar, atmosphere and scene conditions required for desired recognition performance.

The remainder of this thesis is organized as follows. Chapter 2 describes maximum-likelihood (ML) range profile estimation via the expectation-maximization (EM) algorithm. Both planar range profiling approach, which is appropriate for estimating a planar background, and its extension to parametric range profiling, which is used to estimate range data corresponding to arbitrary scenes, are presented. In Chapter 3, the theoretical framework for the model-based statistical object recognition system used in this thesis is explained. The components of the statistical formulation are constructed and used with statistical estimation methods to develop the required object recognition algorithm. Chapter 4 presents the steps followed in the overall object recognition system. The characteristics of the inputs to the system are described and the system modifications needed to perform object recognition on laser radar range imagery are discussed. Chapter 5 concentrates on processing real laser radar range images to extract compact information to be used in the pose estimation and object classification steps. Chapter 6 focuses on estimating the position of the object, as well as identifying the target in the image. The results of our recognition experiments will be presented and analyzed. In Chapter 7, the major conclusions of this work are summarized.

Chapter 2

Maximum-Likelihood Laser Radar Range Imaging

There has been a long interest in the statistics of peak detecting coherent laser radars, which started by examining the fundamentals of single pixel statistics [6] leading to target detection studies for 2-D imagers [7]-[9] and detection [10], [11], planar range profiling [1] and parametric range profiling [2]-[4] for 3-D imagers. Combining theory, experiments and computer simulations, a considerable degree of understanding about the characteristics of laser radar has been achieved, permitting statistical detection and estimation theory for optimum use of sensor data.

These studies lead to an effective algorithm for suppressing noise and range anomalies without appreciable loss of information in the range image. This algorithm will be used as the front-end processor of our model-based object recognition system. Moreover, these techniques will also be used in constructing some of the building blocks of the system. This chapter presents the general theory behind maximum-likelihood laser radar range imaging. Understanding this theory is crucial in using these ideas to accomplish different goals throughout the system.

We start with describing the single-pixel statistical model for the laser radar. A framework for maximum-likelihood (ML) range profiling is presented for fitting a planar

surface to laser radar range data via a computationally convenient approach based on the expectation-maximization (EM) algorithm. Then, the extension of this work to parametric range profiling used to fit a multiresolution basis to an arbitrary scene is discussed. The EM algorithm used in parametric range profiling is modified to yield a computationally efficient and a numerically robust procedure in processing much larger range imagery at high resolutions.

2.1 Measurement Models

A coherent laser radar transmits a series of laser pulses, one for each pixel in a raster scan. The reflected light for each pixel then undergoes optical heterodyne detection, followed by IF (intermediate frequency) filtering, video and peak detection [12], [13], as seen in Fig. 2-1. The range image of some field of view is formed by measuring the time delay between the peaks of the transmitted and detected waveforms. Laser radar range images are degraded by the combined effects of laser speckle and local oscillator shot noise. The former is due to the rough-surfaced nature of the encountered objects when compared to the laser wavelength, which causes constructive and destructive interference in the reflected light [14]. The latter is the fundamental noise encountered in optical heterodyne detection [15] and results in Gaussian noise in the local accuracy of range measurements. Speckle degrades range imagery through range anomalies, which occur when a deep speckle fade combines with a strong noise peak, resulting in a range measurement substantially different from true range value [6], as shown in Fig. 2-2.

Collectively, these degradation mechanisms suggest a statistical approach to laser radar image processing. A statistical characterization for a single pixel of the laser radar data has been theoretically developed and experimentally verified [6]. It takes the form of a conditional probability density function that a measured range value, $r = R$ occurs, given that the true range value is $r^* = R^*$,

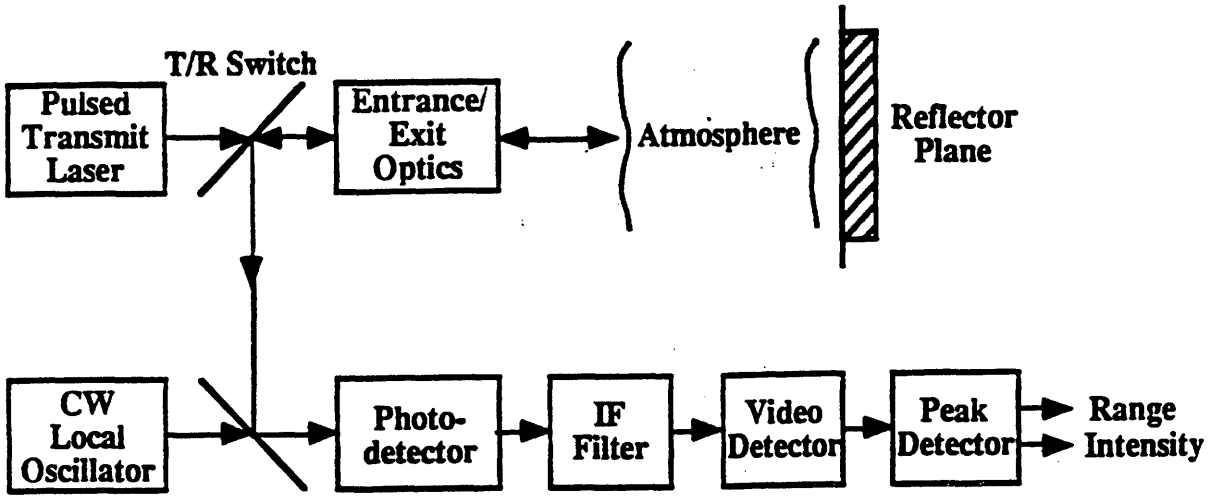

Figure 2-1: Block diagram of a monostatic, shared-optics coherent laser radar.

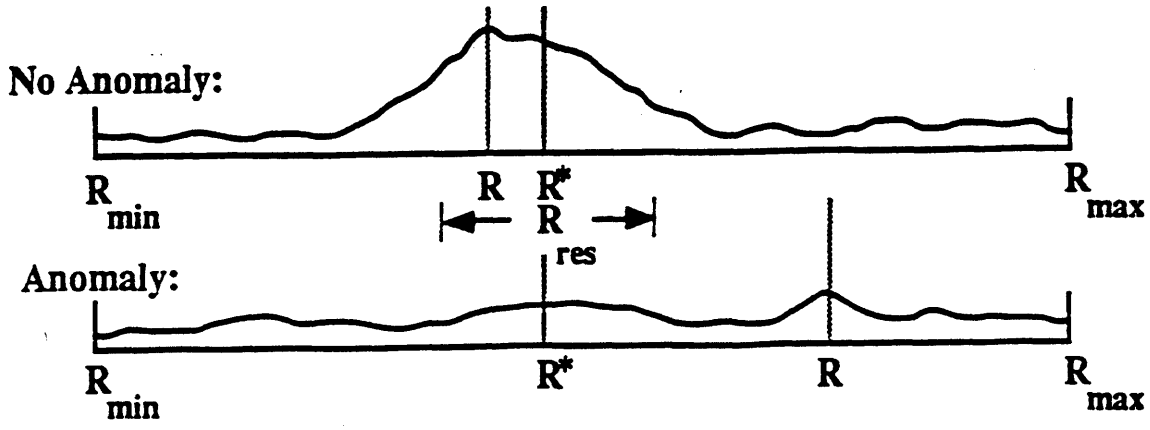

Figure 2-2: Range measurement examples showing anomalous and non-anomalous behaviour.

$$p_{r|r^*}(R|R^*) = [1 - \Pr(A)]\frac{\exp(-\frac{(R-R^*)^2}{2\delta R^2})}{\sqrt{2\pi\delta R^2}} + \Pr(A)\frac{1}{\Delta R} \tag{2.1}$$

In this equation, $\Pr(A)$ is the probability of anomaly, i.e., probability that speckle and shot noise effects combine to yield a range measurement more than one range resolution cell from true range; ΔR is the width of the radar's range uncertainty interval R $\equiv$ [Rmin , Rmax]; and δR is the local range accuracy, i.e., the root-mean-square (rms) range error given the data is not anomalous.

The first term, which is equal to the product of probability that the measurement is not anomalous and a Gaussian probability density with mean equal to the true range value, represents the local range behavior. The second term represents the global range behavior and is equal to the probability that the pixel is anomalous times a uniform distribution of the anomalous range values over the entire range uncertainty interval.

In terms of radar's range resolution, $R_{res} \approx$ cT/2 for a laser pulse with duration T, where c is the speed of light; number of range-resolution bins N $\equiv \Delta R/R_{res}$; and carrier-to-noise ratio,

$$CNR \equiv \frac{\text{average radar return power}}{\text{average local-oscillator shot noise power}} \tag{2.2}$$

the local range accuracy and probability of anomaly are given by

$$\delta R \approx \frac{R_{res}}{\sqrt{CNR}} \tag{2.3}$$

and

$$\Pr(A) \approx \frac{1}{CNR}(\ln(N) - \frac{1}{N} + 0.577), \ for\ CNR \gg 1 \text{ and } N \gg 1 \tag{2.4}$$

By means of Eqs. 2.2 to 2.4, the results of the range estimation problem can be connected to the physical parameters of a real laser radar system. However, the following formulation is confined to the parameters given in Eq. 2.1.

The measured data is a JxK pixel 3-D range image, $\{r_{jk} : 1 \leq j \leq J, 1 \leq k \leq K\}$, where the value of r_{jk} represents the depth of the pixel. The corresponding true range values are, $\{r^*_{jk} : 1 \leq j \leq J, 1 \leq k \leq K\}$. Both the range data and the true range values are rearranged into Q=JK-dimensional column vectors, $\mathbf{r} = \{r_q : 1 \leq q \leq Q\}$ and $\mathbf{r}^* = \{r^*_q : 1 \leq q \leq Q\}$ respectively. For a Q-pixel range image of some field of view, the pixel spacing is usually large enough so that the range measurements are statistically independent given their respective range truths. Thus, the joint probability density that $\mathbf{r} = \mathbf{R}$ given $\mathbf{r}^* = \mathbf{R}^*$, is given by the individual products of single-pixel pdfs,

$$p_{\mathbf{r}|\mathbf{r}^*}(\mathbf{R}|\mathbf{R}^*) = \prod_{q=1}^{Q} \left[[1 - \Pr(A)] \frac{\exp(-\frac{(R_q - R^*_q)^2}{2\delta R^2})}{\sqrt{2\pi\delta R^2}} + \Pr(A) \frac{1}{\Delta R} \right] \tag{2.5}$$

The laser radar range profiling problem is then to find the optimal range estimate, given this likelihood function. To find the optimal range estimate of the true range image, $\mathbf{r}^*$, given the measured data, $\mathbf{r}$, maximum-likelihood (ML) estimation is employed. Given a particular observation vector, $\mathbf{R}$, the ML estimate of the range data is the $\mathbf{R}^*$ that maximizes $p_{\mathbf{r}|\mathbf{r}^*}(\mathbf{R}|\mathbf{R}^*)$, i.e., it maximizes the likelihood of our observing the data vector, $\mathbf{R}$, we have obtained.

$$\widehat{\mathbf{r}}^*_{ML}(\mathbf{R}) = \arg\max_{R^*}(p_{\mathbf{r}|\mathbf{r}^*}(\mathbf{R}|\mathbf{R}^*)) \tag{2.6}$$

However, the joint probability density in Eq. 2.5 implies that the ML estimate of the range image is the raw data itself. This means that the range anomalies, which may occur on more than 10% of the pixels at reasonable CNR's, cannot be suppressed by this method. Therefore, an additional resolution constraint is provided in the problem to serve the purpose of suppressing the anomalies in the range data while at the same time giving the desired resolution to image features.

2.2 Planar Range Profile Estimation

The objective of planar range profiling is to find the optimal estimate of the true range image, $\mathbf{r}^*$, given the observed data, $\mathbf{r}$. The pixel values of the true range are assumed to comprise a plane, given by

$$r^*_{jk} = x_1 j + x_2 k + x_3 \qquad 1 \leq j \leq J \ , \ 1 \leq k \leq K \tag{2.7}$$

where x_1 and x_2 are the elevation and azimuth range slopes, respectively, and x_3 is the range intercept. These three parameters are to be estimated using the measured range data. The 3-D parameter vector that characterizes the planar profile will be defined as

$$\mathbf{x} = [x_1 \ x_2 \ x_3]^T \tag{2.8}$$

Similar to Eq. 2.5, we can express the joint probability density for $\mathbf{r} = \mathbf{R}$ to occur, given $\mathbf{x} = \mathbf{X}$ as

$$p_{\mathbf{r}|\mathbf{x}}(\mathbf{R}|\mathbf{X}) = \prod_{j=1}^{J} \prod_{k=1}^{K} \left[[1 - \Pr(A)] \frac{\exp(-\frac{(R_{jk} - R^*_{jk})^2}{2\delta R^2})}{\sqrt{2\pi\delta R^2}} + \Pr(A) \frac{1}{\Delta R} \right] \tag{2.9}$$

To estimate the parameter vector $\mathbf{x}$, maximum-likelihood estimation is employed. Since the logarithm is a monotonically increasing function, the logarithm of the likelihood function can be maximized to obtain the maximum-likelihood estimate, $\widehat{\mathbf{x}}_{ML}$. The necessary condition that needs to be satisfied for an extremum at $\mathbf{X} = \widehat{\mathbf{x}}_{ML}$ is

$$\frac{\partial}{\partial \mathbf{X}} \ln[p_{\mathbf{r}|\mathbf{x}}(\mathbf{R}|\mathbf{X})]|_{\mathbf{X}=\widehat{\mathbf{x}}_{ML}} = 0 \tag{2.10}$$

Plugging Eq. 2.9 into Eq. 2.10 leads to a nonlinear vector equation for $\widehat{\mathbf{x}}_{ML}$,

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\begin{bmatrix} j \\ k \\ 1 \end{bmatrix}[R_{jk}-(X_1 j+X_2 k+X_3)]\times w_{jk}(\mathbf{X})\Big|_{\mathbf{X}=\hat{\mathbf{x}}_{ML}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.11}$$

where X_1, X_2 and X_3 are the components of the vector $\mathbf{X}$ and the jkth pixel weight, w_{jk}, is defined as

$$w_{jk}(\mathbf{X}) = \frac{[1-\Pr(A)]\frac{\exp\left(-\frac{R_{jk}-R^*_{jk}(\mathbf{X})^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}}{[1-\Pr(A)]\frac{\exp\left(-\frac{\left(R_{jk}-R^*_{jk}(\mathbf{X})\right)^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}+\Pr(A)\frac{1}{\Delta R}} \tag{2.12}$$

where

$$R^*_{jk}(\mathbf{X}) = X_1 j + X_2 k + X_3 \tag{2.13}$$

From Eq. 2.12, it is clear that we have $0 \leq w_{jk} \leq 1$, i.e., the weights of the pixels are proper fractions. Indeed, $w_{jk}(\mathbf{X})$ represents the conditional probability that r_{jk} is not anomalous, assuming that the true parameter vector is $\mathbf{X}$. Eq. 2.11 appears to be a linear equation in $\mathbf{X}$, but it is not since the weights w_{jk}, are functions of $\mathbf{X}$. If the anomaly probability is very small, so that $w_{jk} \approx 1$ for all j, k, then Eq. 2.11 becomes linear and its solution can easily be shown to be,

$$\hat{\mathbf{x}}_{ML} = \mathbf{G}^{-1}\mathbf{K} \tag{2.14}$$

where the matrices $\mathbf{G}$ and $\mathbf{K}$ are defined as

$$\mathbf{G} = \sum_{j=1}^{J}\sum_{k=1}^{K}\begin{bmatrix} j \\ k \\ 1 \end{bmatrix}\begin{bmatrix} j & k & 1 \end{bmatrix} \tag{2.15}$$

and

$$\mathbf{K} = \sum_{j=1}^{J} \sum_{k=1}^{K} \begin{bmatrix} j \\ k \\ 1 \end{bmatrix} R_{jk} \tag{2.16}$$

However, the probability of anomaly is usually substantial in most of the data obtained by the laser radar. Therefore, the nonlinear nature of the of the problem cannot be ignored. The nonlinear estimation problem will be solved by an iterative approach using the expectation-maximization algorithm, which will be presented in the next section.

2.2.1 Expectation-Maximization Algorithm

The EM algorithm is effectively used for ML estimation problems in which the observation vector constitutes incomplete data [16]. An incomplete data problem is one in which the observation vector, available for processing, is only a part of the complete data vector and there is a degree of freedom for constructing a complete data vector. In ML range profiling, the natural complete data vector is

$$\mathbf{y} = \begin{bmatrix} \mathbf{r} \\ \mathbf{a} \end{bmatrix} \tag{2.17}$$

where $\mathbf{r}$ is the range observation vector and $\mathbf{a}$ is the anomaly data, which is the missing part of the complete data. If the complete data vector were available, the anomalous pixels would be identified and suppressed, and the ML range profiling problem would become a linear problem including nonanomalous pixels only. However, since $\mathbf{a}$ is not directly observed, this perfect elimination is not possible, and the ML estimation tries to deal with the possibility of anomalous pixels in a statistical fashion, resulting in a nonlinear estimation problem. The iterative EM algorithm is a computationally simple procedure to solve this problem because of the linear nature of the complete data problem.

The EM algorithm starts from an initial estimate of the parameter vector, $\widehat{\mathbf{x}}(0)$, and

produces a sequence of parameter estimates, $\{\ \widehat{\mathbf{x}}(n) : n = 1, 2, 3, ...\}$ by an alternating sequence of expectation and maximization steps. The associated likelihood sequence is monotonically increasing. Hence, the EM algorithm converges to a likelihood maximum.

However, the EM algorithm, being a local nonlinear optimization method, needs a good initial starting value to converge to the correct local maximum. If the initial estimate is good enough to place the EM algorithm on the highest hill, the global maximum will be achieved. For imagery with low probability of anomaly, $Pr(A) \leq 0.1$, a linear least-squares (LS) initial estimate is sufficient. For many cases of interest, however, such a simple initial estimate is insufficient for reliable location of the global likelihood maximum. The recursive expectation-maximization algorithm is presented in the next section as an extension of the LS-initialized EM estimation procedure.

In essence, the LS-initialized EM algorithm starts by assuming that all of the pixels are nonanomalous, which corresponds to solving Eq. 2.11 assuming $w_{jk} = 1$ for all j, k. The EM algorithm uses the latest estimate to update the weights and then solves the linear estimation problem with the new weights, treating them as constants. In particular, when $\{w_{jk}(n),\ \widehat{\mathbf{x}}(n)\}$ pair is available, $\{w_{jk}(n+1),\ \widehat{\mathbf{x}}(n+1)\}$ is obtained by a two-step procedure:

1. First the **expectation** step updates the weights by means of

$$w_{jk}(n+1) = \frac{[1 - \Pr(A)]\dfrac{\exp\left(-\frac{\left(R_{jk} - R^*_{jk}(\widehat{\mathbf{x}}(n))\right)^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}}{[1 - \Pr(A)]\dfrac{\exp\left(-\frac{\left(R_{jk} - R^*_{jk}(\widehat{\mathbf{x}}(n))\right)^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}} + \Pr(A)\frac{1}{\Delta R}} \tag{2.18}$$

2. Next, the **maximization** step updates the estimate by means of

$$\widehat{\mathbf{x}}_{ML}(n+1) = \mathbf{G}(n+1)^{-1}\mathbf{K}(n+1) \tag{2.19}$$

where the matrices $\mathbf{G}(n+1)$ and $\mathbf{K}(n+1)$ are defined as

$$\mathbf{G}(n+1)=\sum_{j=1}^{J}\sum_{k=1}^{K}\begin{bmatrix} j \\ k \\ 1 \end{bmatrix} w_{jk}(n+1)\begin{bmatrix} j & k & 1 \end{bmatrix} \tag{2.20}$$

and

$$\mathbf{K}(n+1)=\sum_{j=1}^{J}\sum_{k=1}^{K}\begin{bmatrix} j \\ k \\ 1 \end{bmatrix} w_{jk}(n+1)R_{jk} \tag{2.21}$$

Since the likelihood function is increasing in each step, the EM algorithm is guaranteed to provide improving parameter estimates. This iterative process is terminated when the difference between successive likelihoods lies within some predetermined threshold. The final estimate is the ML estimate if the initial estimate is on the highest likelihood hill.

2.2.2 Recursive EM Algorithm

The Recursive EM (REM) algorithm is an extension of the least-squares (LS) initialized EM estimation procedure. It has been shown to provide better initialization, using a recursive approach, for cases with appreciable anomaly probabilities, both in planar [1] and parametric [4] range profiling work.

The REM Algorithm begins by setting the local range accuracy, δR, in the single pixel range density equal to the whole range uncertainty interval, ΔR. The resulting density is used in an LS initialized EM algorithm to obtain the zeroth order estimate. Then the local range accuracy is set to half of the range uncertainty interval, i.e., half of the value for the local range accuracy used in the previous recursion. This time the resulting density is used in $\widehat{\mathbf{x}}_{rem}(0)$ initialized EM algorithm. This process goes on until the local range accuracy is set to δR. The output of the last stage is the final REM estimate.

Because the local range accuracy is gradually decreased in the REM algorithm, only

the most likely anomalous pixels are discarded. This reduces the chance that a lot of non-anomalous pixels are removed from the estimate, making it more likely for the REM estimate to coincide with the ML estimate, because of the quality of the initial estimate.

2.2.3 Range Profile Results

In this section, we will demonstrate the results for planar range profiling of simulated laser radar range imagery. The sample image to be profiled corresponds to a planar background without any target. The elevation/azimuth angles and the range intercept of this plane are selected to be consistent with those of the planar background associated with the real laser radar range image taken from the available data release used later. The generated image is 45×128, which is the size of the range images in the data set. This image can be seen in Fig. 2-3. In this figure, the gray level of each pixel represents the distance in range bins where one range bin is equal to 1.1 meters. The imaged field-of-view is about 400 to 500 range bins away from the laser radar as shown by the calibration bar on the right. Considering the additional range gate offset of 427 meters, the actual distance is approximately 870 to 980 meters. For this particular image almost no anomalies can be observed. Thus, in order to test the EM/ML procedure, range data is synthesized using this image as the range truth by adding noise and simulating the anomalies. In particular, zero-mean Gaussian noise with a standard deviation, $\delta R = 2$, is added to each pixel of the range truth and anomalies are simulated at a 5% rate using a uniform probability density across the range uncertainty interval, $\Delta R = 1524$ bins. This process will be discussed in more detail in Chapter 4. The resulting range image is shown in Fig. 2-4.

A planar range surface is fitted to this range data using the procedure described earlier. The fitted planar background is shown in Fig. 2-5.

The resulting profile suppresses the anomalies, which appear as black and white pixels in the range data in Fig. 2-4. The algorithm assumes that the range data constitute a plane. Therefore, planar range profiling algorithm is most effective for a very restricted

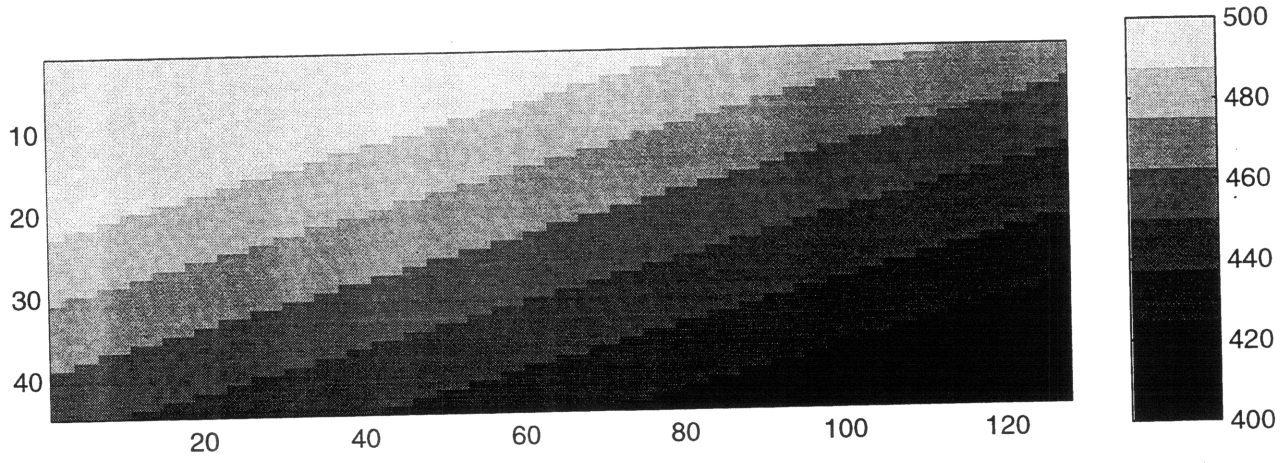

Figure 2-3: Range image of a planar surface.

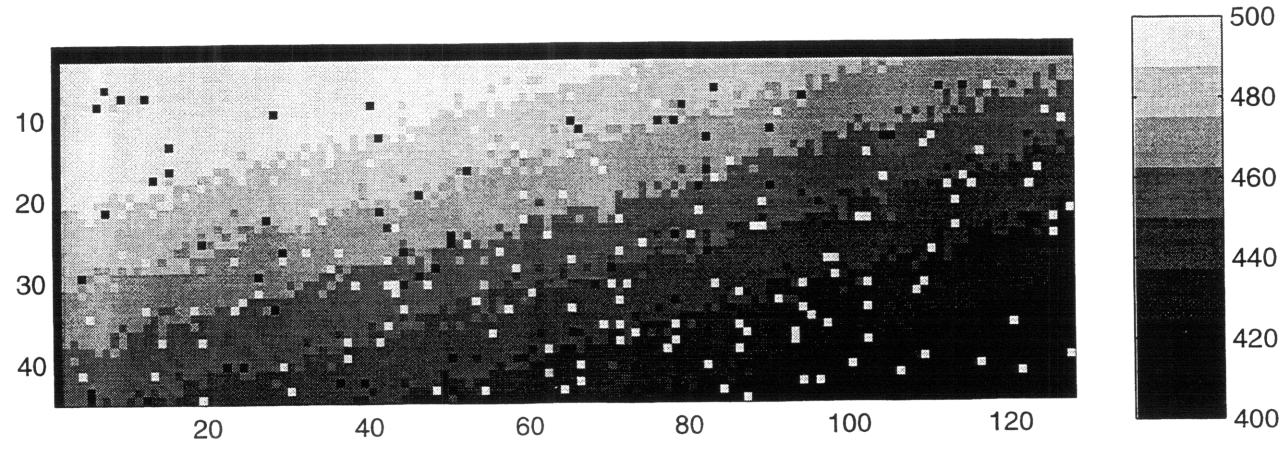

Figure 2-4: Range data of a planar surface, artificially created from the range truth by addition of statistically independent, zero mean Gaussian noise to each pixel and random creation of anomalies.

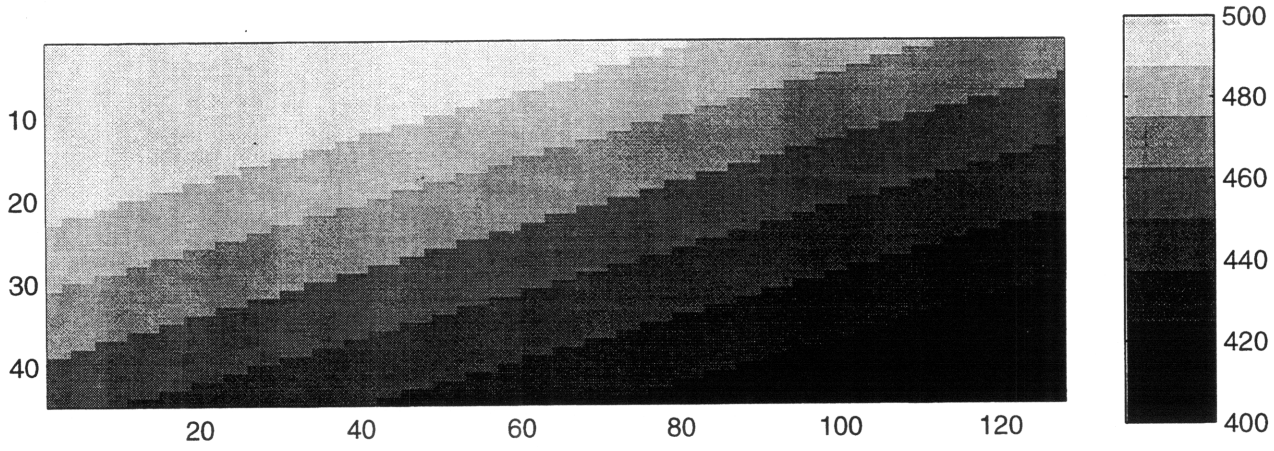

Figure 2-5: Planar range profile fitted to the planar surface.

type of range data, having a planar terrain without any target. To process range imagery in which a target is present, this approach should be applied to the target region and the background region separately. In the next section the planar range profiling work is extended to parametric range profiling, which involves fitting a multiresolution wavelet basis to any type of scenery.

2.3 Parametric Range Profile Estimation

Range profiling using the ML estimation approach need not be confined to fitting a planar surface to range data. In this section the framework for the more general case of parametric range profiling based on the EM algorithm is briefly presented.

For this method, we need to derive a parametric representation for the true range vector, $\mathbf{r}^*$, of length Q,

$$\mathbf{r}^* = \begin{bmatrix} r_1^* \\ \vdots \\ r_Q^* \end{bmatrix} \tag{2.22}$$

We denote the parameter vector by $\mathbf{x}$, which is also of length Q,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_Q \end{bmatrix} \tag{2.23}$$

and an associated orthogonal $Q \times Q$ transformation matrix by $\mathbf{H}$, whose columns, $\{\Phi_q : 1 \leq q \leq Q\}$, form an orthonormal basis for the Q-length vector space,

$$\mathbf{H} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_Q \end{bmatrix} \tag{2.24}$$

This transformation matrix is defined to be such that it transforms the true range vector into a parameter vector via

$$\mathbf{x} = \mathbf{H}^T \mathbf{r}^* \tag{2.25}$$

Since $\mathbf{H}$ is an orthogonal matrix, $\mathbf{H}^{-1} = \mathbf{H}^T$, the true range vector can be represented as

$$\mathbf{r}^* = \mathbf{H}\mathbf{x} \tag{2.26}$$

Suppose, now that the true range can be characterized by a parameter vector, $\mathbf{x}$ of length $P < Q$, i.e., only the first P dimensions of $\mathbf{x}$ are non-zero, then,

$$\mathbf{r}^* = \mathbf{H}_P \mathbf{x}_P \tag{2.27}$$

where

$$\mathbf{x}_P = \begin{bmatrix} x_1 \\ \vdots \\ x_P \end{bmatrix} \tag{2.28}$$

and

$$\mathbf{H}_P = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_P \end{bmatrix} \tag{2.29}$$

The last $Q - P$ columns in $\mathbf{H}$ are then not used in the characterization of the range truth and hence not used in finding the range estimate. This provides a means of selecting the resolution, P, of the estimated range data since choosing $P < Q$ involves suppressing certain data, which is beneficial, since our objective is to suppress anomalous data.

The likelihood function is now conditioned on $\mathbf{x}_P$, the parameter vector, with $\mathbf{R}^*$ replaced by $\mathbf{H}_P\mathbf{x}_P$ and is given by

$$p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R}|\mathbf{X}_P) = \prod_{q=1}^{Q} \left[[1 - \Pr(A)] \frac{\exp(-\frac{(R_q - (\mathbf{H}_P\mathbf{X}_P)_q)^2}{2\delta R^2})}{\sqrt{2\pi\delta R^2}} + \Pr(A)\frac{1}{\Delta R} \right] \tag{2.30}$$

where $R_q^* = (\mathbf{H}_P\mathbf{X}_P)_q$ is the qth component of the true range vector $\mathbf{R}^*$. Note that estimating $\mathbf{r}^*$ from $\mathbf{r}$ is equivalent to estimating $\mathbf{x}_P$ from $\mathbf{r}$,because

$$\widehat{\mathbf{r}}_{ML}^*(\mathbf{R}) = \mathbf{H}_P\widehat{\mathbf{x}}_{PML}(\mathbf{R}) \tag{2.31}$$

The rest of the procedure is very similar to planar range profiling. ML estimation is used to estimate the true range vector, $\mathbf{r}^*$, given the observed range vector, $\mathbf{r}$. The necessary condition that needs to be satisfied by the ML estimate is

$$\frac{\partial}{\partial \mathbf{X}_P} \ln[p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R}|\mathbf{X}_P)]|_{\mathbf{x}_P = \widehat{\mathbf{x}}_{PML}} = \mathbf{0} \tag{2.32}$$

for an extremum point. Substituting Eq. 2.30 into Eq. 2.32 yields a nonlinear vector equation,

$$\frac{1}{\delta R^2}\mathbf{H}_P^T\mathbf{W}(\mathbf{X}_P)(\mathbf{R}-\mathbf{H}_P\mathbf{X}_P)|_{\mathbf{X}_P=\widehat{\mathbf{x}}_{PML}}=\mathbf{0} \tag{2.33}$$

where $\mathbf{W}(\mathbf{X}_P)$ is a $Q\times Q$ diagonal matrix. The qqth element of $\mathbf{W}$ is the qth weight w_q, which is equal to the conditional probability that the associated pixel is not anomalous, given that the true parameter vector is $\mathbf{X}_P$,

$$w_q(\mathbf{X}_P)=\frac{[1-\Pr(A)]\frac{\exp\left(-\frac{(R_q-(\mathbf{H}_P\mathbf{X}_P)_q)^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}}{[1-\Pr(A)]\frac{\exp\left(-\frac{(R_q-(\mathbf{H}_P\mathbf{X}_P)_q)^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}+\Pr(A)\frac{1}{\Delta R}} \tag{2.34}$$

If probability of anomaly is very small then Eq. 2.33 becomes linear and it is easily solved. In general, however, the probability of anomaly cannot be neglected. We solve the nonlinear estimation problem iteratively via the EM algorithm, as explained in Section 2.1.2. In particular, the algorithm calculates the weights by Eq. 2.34, using the latest estimate for the parameter vector and then uses the recently calculated weights to estimate the new parameter vector, via Eq. 2.33. The REM algorithm is used to solve the initialization problem for the EM algorithm.

2.3.1 Haar Wavelet Basis & Fast EM-ML Algorithm:

Parametric range imaging is employed so that we can impose regularity conditions, which ensure a certain degree of anomaly suppression. If the true range image may be assumed to be reasonably planar, the natural parametric model to employ is given in planar range profiling work in Eq. 2.7. However, for more general type of imagery, as when we are trying to profile the target and the background simultaneously in the image, the parametric model to be used is not so clear. The natural approach to follow in this case is to use a wavelet basis, $\{\Phi_q\}$, which has been ordered such that increasing q corresponds to increasingly fine scale behavior.. In this way, it is possible to extract coarse-scale

features from the range data first, by using a small P value, and then to progress to find estimates of increasing resolution by increasing P. A weight-based procedure, using the statistics of the number of anomalous, low-weighted, pixels, for optimally terminating the coarse-to-fine scale progression of EM/ML range imaging is given in [3].

In the previous work [2]-[4], the Haar wavelet basis was used to construct the orthogonal transformation matrix. This transformation matrix is formed of orthonormal column vectors, Φ_q, such that increasing q corresponds to increasing resolution, as required. This multiresolution nature of the Haar wavelet basis permits ML range imaging at any desired resolution.

To obtain some initial understanding for the choice of this basis, we focus on 1-D Haar wavelet basis first. The extension from 1-D to 2-D Haar wavelet basis, which is used to profile the 3-D laser radar range imagery, is straightforward.

In 1-D Haar wavelet basis, the wavelets have a progression of finer scale behavior. Below is an example to illustrate the nature of the Haar wavelet basis for P=8 and Q=8:

$$\Phi_1^{\mathrm{T}} = \sqrt{\tfrac{1}{8}}\,[1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$
$$\Phi_2^{\mathrm{T}} = \sqrt{\tfrac{1}{8}}\,[1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1]$$
$$\Phi_3^{\mathrm{T}} = \sqrt{\tfrac{2}{8}}\,[1 \quad 1 \quad -1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0]$$
$$\Phi_4^{\mathrm{T}} = \sqrt{\tfrac{2}{8}}\,[0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad -1 \quad -1]$$
$$\Phi_5^{\mathrm{T}} = \sqrt{\tfrac{4}{8}}\,[1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$
$$\Phi_6^{\mathrm{T}} = \sqrt{\tfrac{4}{8}}\,[0 \quad 0 \quad 1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0]$$
$$\Phi_7^{\mathrm{T}} = \sqrt{\tfrac{4}{8}}\,[0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -1 \quad 0 \quad 0]$$
$$\Phi_8^{\mathrm{T}} = \sqrt{\tfrac{4}{8}}\,[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad -1]$$

The range estimate is composed of some linear combination of the wavelets $\{\Phi_p\}$. By choosing P, a piecewise constant profile at a particular resolution can be fitted to the range data. In fact the range estimate, for P a power of two, consists of P piecewise constant Q/P length intervals .

The 2-D Haar wavelet basis, used with real laser radar imagery, is constructed by multiplying two 1-D bases, one in each of the image dimensions, to form the 2-D Haar

wavelet basis. Suppose $\mathbf{H}_{P_j}$, which is $J \times P_j$, and $\mathbf{H}_{P_k}$, which is $K \times P_k$, are 2 initial 1-D bases,

$$\mathbf{H}_{P_j} \equiv \left[\begin{array}{ccc} \Phi_1 & \cdots & \Phi_{P_j} \end{array} \right] \text{ where } \Phi_j = \left[\begin{array}{c} \phi_{j1} \\ \vdots \\ \phi_{jJ} \end{array} \right] \text{ , for } 1 \leq j \leq P_j \tag{2.35}$$

and

$$\mathbf{H}_{P_k} \equiv \left[\begin{array}{ccc} \Phi_1 & \cdots & \Phi_{P_k} \end{array} \right] \text{ where } \Phi_k = \left[\begin{array}{c} \phi_{k1} \\ \vdots \\ \phi_{kK} \end{array} \right] \text{ , for } 1 \leq k \leq P_k \tag{2.36}$$

The 2-D Haar wavelet basis, $\mathbf{H}_{P_jP_k}$, is given by,

$$\mathbf{H}_{P_jP_k} \equiv \left[\begin{array}{ccc} \Psi_{11} & \cdots & \Psi_{P_jP_k} \end{array} \right] \text{ , for } 1 \leq j \leq P_j \text{ and } 1 \leq k \leq P_k \tag{2.37}$$

where $\{\Psi_{jk}\}$ is the column-vector basis for the $Q = JK$ -length vector space, given by,

$$\Psi_{jk} = \left[\begin{array}{c} \Phi_j\phi_{k1} \\ \Phi_j\phi_{k2} \\ \vdots \\ \Phi_j\phi_{kK} \end{array} \right] \tag{2.38}$$

Notice that it is possible to associate different resolutions, P_j and P_k values, for the two 1-D bases. Thus the range data can be estimated using different resolutions in elevation and azimuth directions. For convenience in notation, we define $P = P_jP_k$, $Q = JK$ and $\Phi_p = \Psi_{jk}$,where $1 \leq p \leq P$, $1 \leq j \leq J$, and $1 \leq k \leq K$. As a result , the $Q \times P$ 2-D Haar wavelet basis, $\mathbf{H}_P$ can be written as;

$$\mathbf{H}_P \equiv \mathbf{H}_{P_jP_k} \equiv \left[\begin{array}{ccc} \Phi_1 & \cdots & \Phi_P \end{array} \right] \tag{2.39}$$

In the range estimation problem, fitting this basis to range data corresponds to esti-

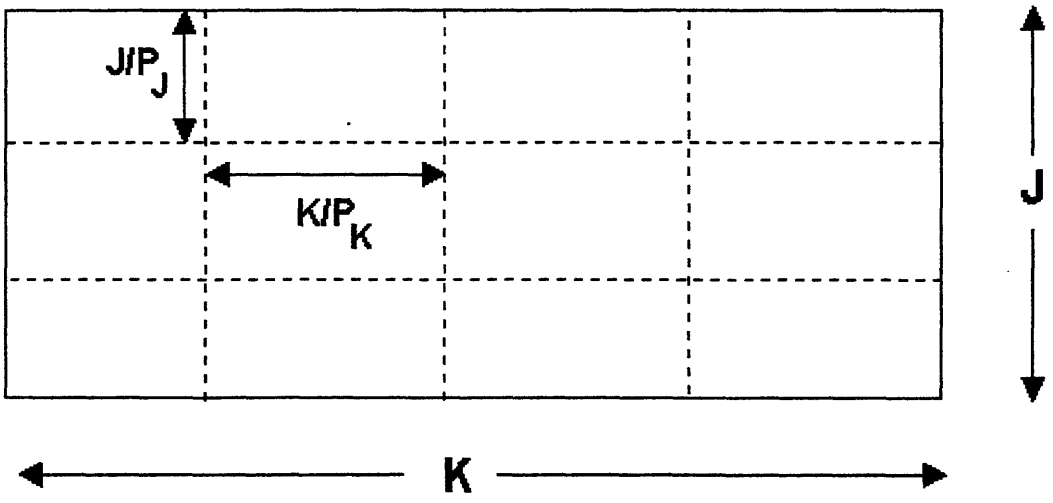

Figure 2-6: Illustration of 2-D Haar wavelet range space.

mating the values of $P = P_j P_k$ blocks of constant value in an image of $Q = JK$ pixels, where each block is J/P_j by K/P_k as illustrated in Fig. 2-6.

The conventional EM algorithm can only be used to range profile small imagery at low resolutions due to calculational complexity of the algorithm. The computational complexity of the conventional EM Algorithm is dominated by the maximization step, which involves a huge matrix inversion and this load increases as the image size Q and the resolution P are increased. However using the special structure of Haar wavelet basis, in particular the non-overlapping support nature of the range space of this basis, the maximization step is converted from a matrix inversion to a simple matrix multiplication [2] , which yields a procedure (the fast EM/ML algorithm) that is both computationally efficient and numerically robust. Essentially, the algorithm splits the Q-pixel range image into P blocks, each of size Q/P pixels, as shown in Fig. 2-6, and estimates the value of each block by the weighted sum of pixel values in the block, normalized by the sum of the weights.

$$\widehat{r_q}(n) = \frac{\sum_{i \in Q_s(p)} w_i(n) R_i}{\sum_{i \in Q_s(p)} w_i(n)} \qquad \text{for } p \text{ such that i} \in Q_s(p) \tag{2.40}$$

where $\{Q_s(p) : 1 \leq p \leq P\}$ represents a nonoverlapping tiling of the image and $w_i(n)$ is the conditional probability that the associated pixel is not anomalous given that the most recent parameter vector estimate is correct.

By means of this algorithm, it is possible to profile much larger imagery at much higher resolutions, at a calculation speed increased by many orders of magnitude.

2.3.2 Range Profile Results

In this section, we present the results for parametric range profiling of real imagery using fast EM/ML algorithm at several resolutions. The image to be profiled contains an armored personnel carrier (apc) as the target located behind a tree and a pole. The video image can be seen in Fig. 2-7. The corresponding range image to be profiled is shown in Fig. 2-8. Note that the field-of view of the laser radar range sampler is more focused than that of the video recorder. The tree and the pole appear partially in the range image as dark pixels which represent low range values with respect to the sensor location. The range image is contained between 780 and 900 range bins or equivalently 1285 and 1420 meters away.

Similar to the range image processed by planar range profiling in Section 2.2.3 , this range image has almost no anomalies, so the actual range data is generated from this image assuming it to be the range truth by adding zero-mean Gaussian noise with standard deviation, $\delta R = 2$, and simulating the anomalies at 5% rate. The resulting range image is shown in Fig. 2-9. The pixel values on the edges are preset to zero to remove the edge effects.

Given the range data and the necessary parameters, range profiling is performed at various resolutions , $P = \{512, 1024, 2048, 4096\}$, to find the maximum-likelihood range estimate by the fast EM/ML approach. This corresponds to using 4×8-pixel, 4×4-pixel,

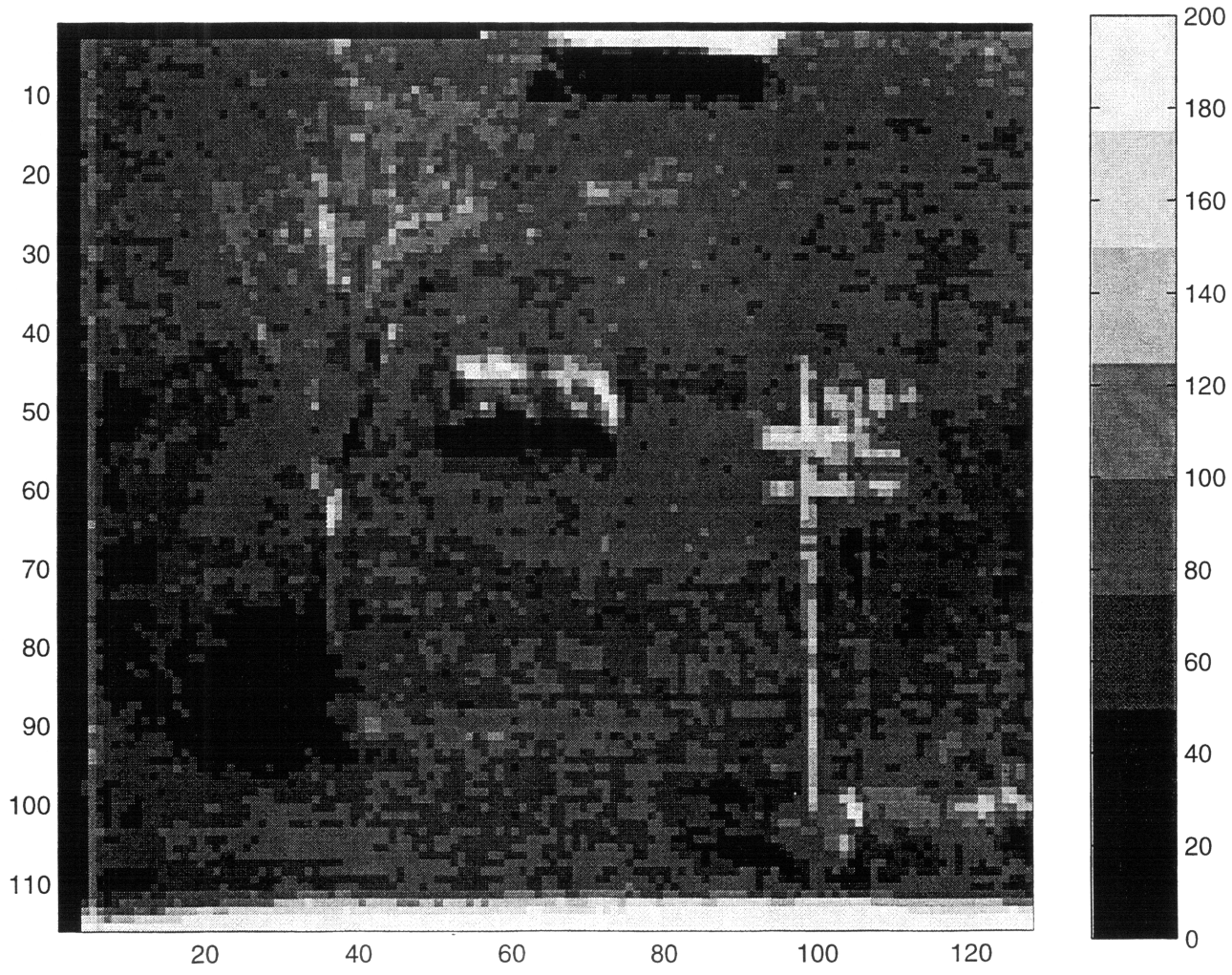

Figure 2-7: Video image of an armored personnel carrier.

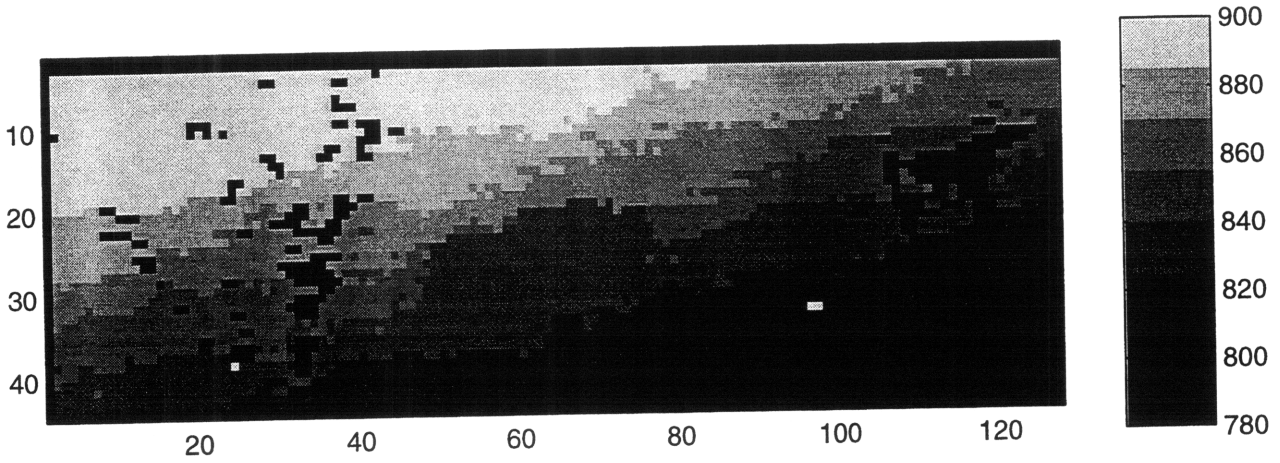

Figure 2-8: Range image of an armored personnel carrier.

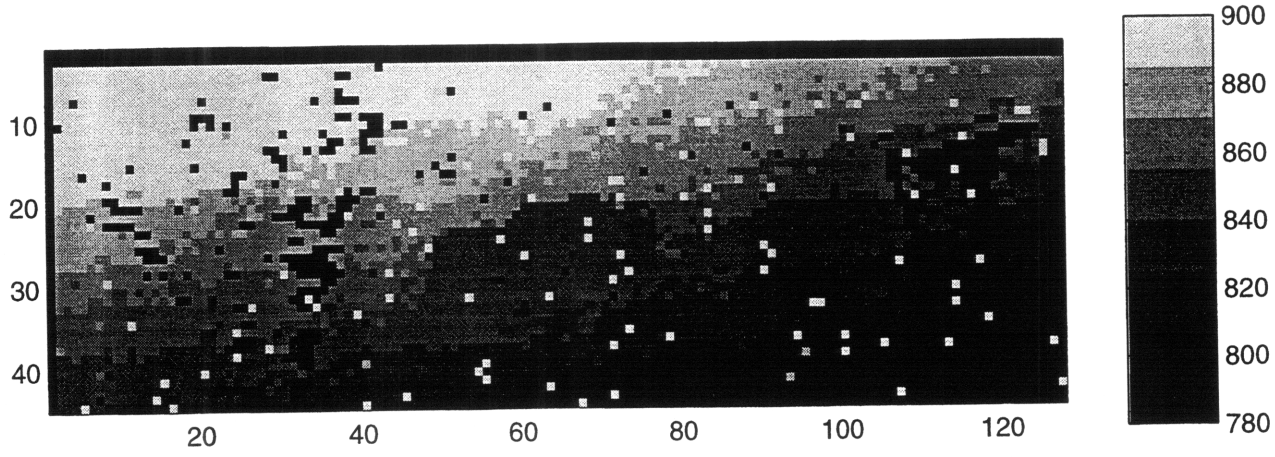

Figure 2-9: Range data of an armored personnel carrier, artificially created from the range truth by addition of statistically independent, zero mean Gaussian noise to each pixel and random creation of anomalies.

2×4-pixel, and 2×2-pixel blocks respectively. The resulting profiles are shown in Figs. 2-10 to 2-13.

As seen in these profiles, the anomalies are almost completely suppressed. In this particular range image, the target apc is located sufficiently far from the laser radar that it is hard to discern the target's shape because of the limited number of pixels on the target. Therefore at low spatial resolutions, the target cannot be located clearly. Since the tree and the pole consist of thin branches, most of the pixels with them are suppressed by the large amount of background pixels in large block sizes. The range profile displays more detailed information about the image as the resolution increases. At the highest resolution, the general outline of the target, the tree and the pole can be well observed against the planar, sloping, but otherwise featureless background.

Some other insight into the operation of the fast EM/ML algorithm is provided by the final weights associated with the profiles, which are arranged into 45×128 pixel images. The weight images are demonstrated in Figs. 2-14 to 2-17. Weight images are specifically useful for this image due to the presence of the tree and the pole, since the pixels that constitute these objects are not concentrated on a region, but appear isolated as if they were anomalous pixels.

Note that the weights for the pixels vary from 0, meaning completely anomalous, to 1, meaning completely non-anomalous. In the range profiles, for each block, pixel values close to the block's range estimate are weighted close to one and those far from this estimate are weighted close to zero. Especially at lower resolutions, the weight images clearly display the boundary between the target and the background. This phenomenon is easy to understand. When the block size is large, the EM algorithm suffers in fitting the wide variation in pixel values inside a block at a range discontinuity. Eventually, the estimate for the block will be close to the range value of the majority of the pixels and the rest of the pixels inside the block will be treated as anomalous and are weighted close to zero, as shown in the weight images. Therefore, the weight images of low resolution fits can be used to locate parts of an object that are significantly smaller in one dimension

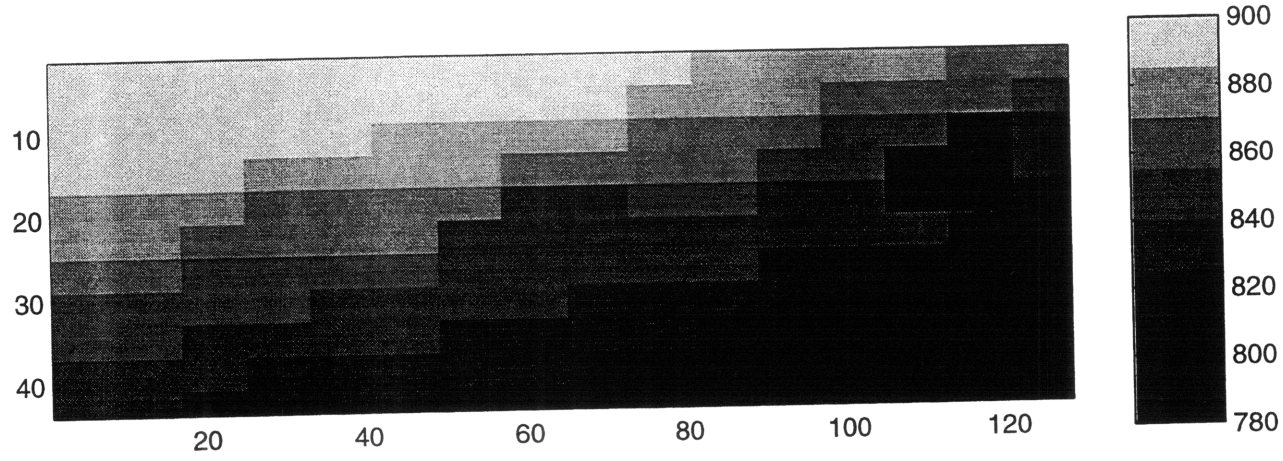

Figure 2-10: Multiresolution Haar wavelet EM/ML 4×8 fit to range data.

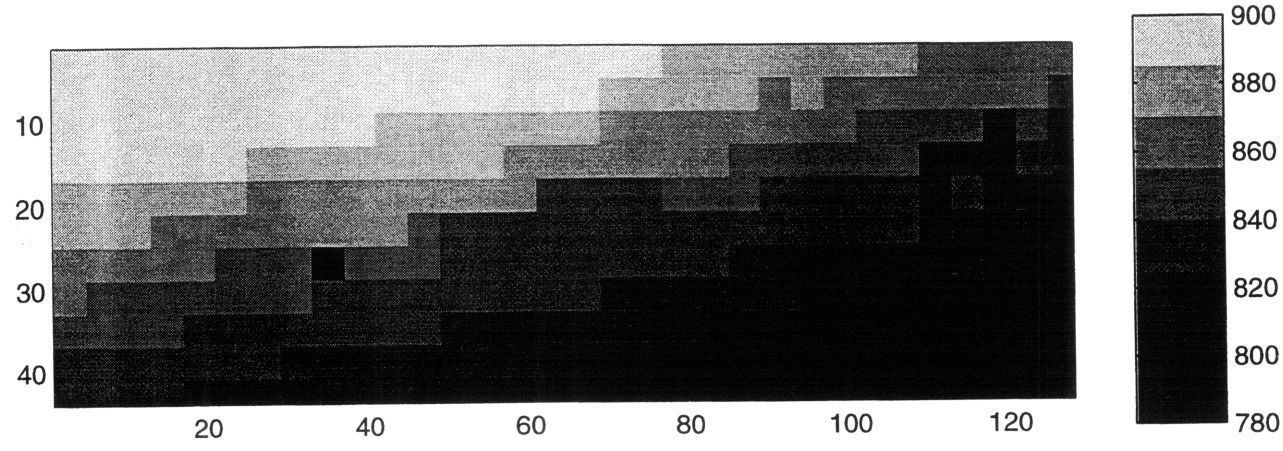

Figure 2-11: Multiresolution Haar wavelet EM/ML 4×4 fit to range data.

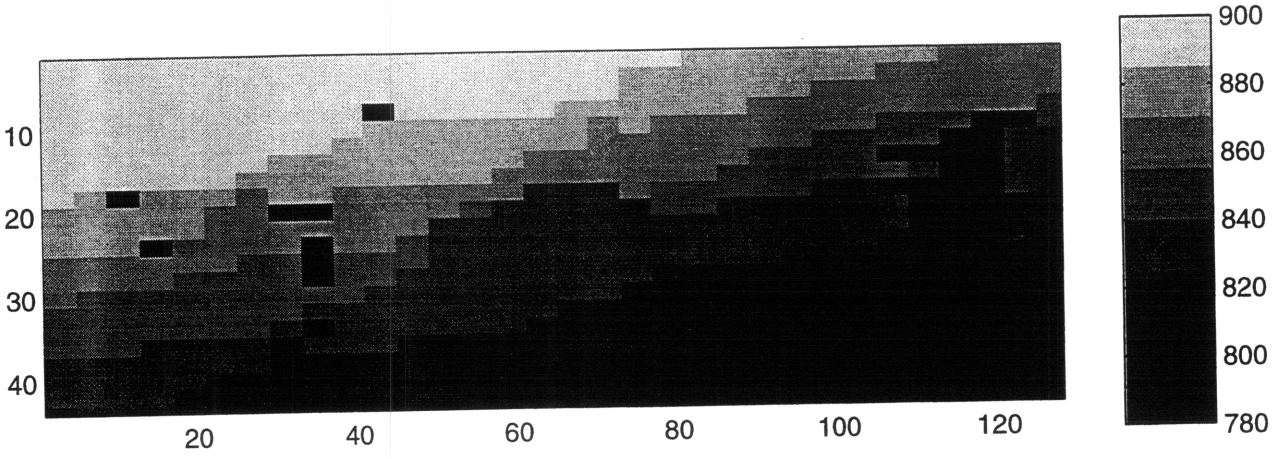

Figure 2-12: Multiresolution Haar wavelet EM/ML 2×4 fit to range data.

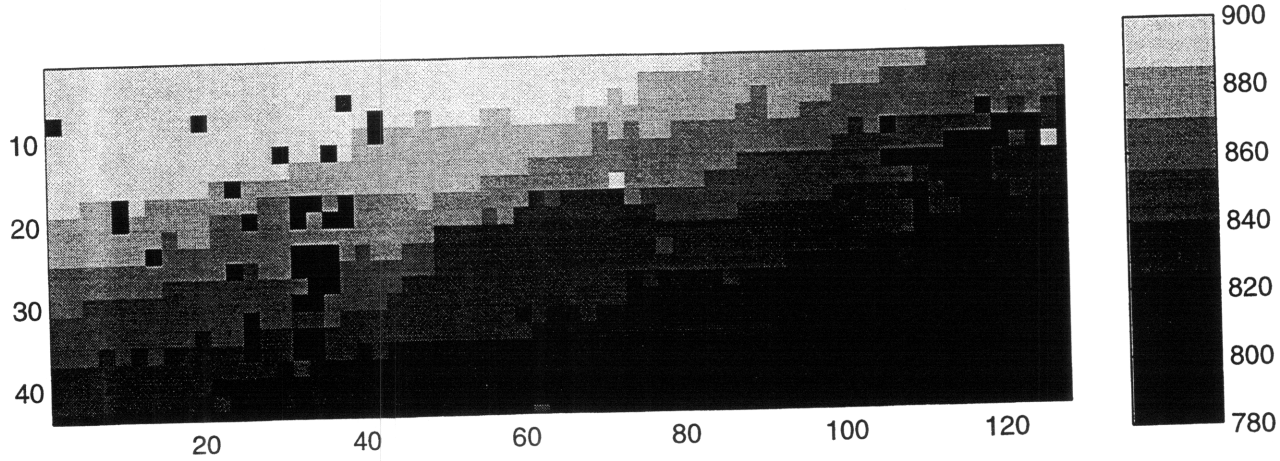

Figure 2-13: Multiresolution Haar wavelet EM/ML 2×2 fit to range data.

than the block size used in the ML/EM fit or to detect the edges of an object that appear as a cluster of pixels in the weight image.

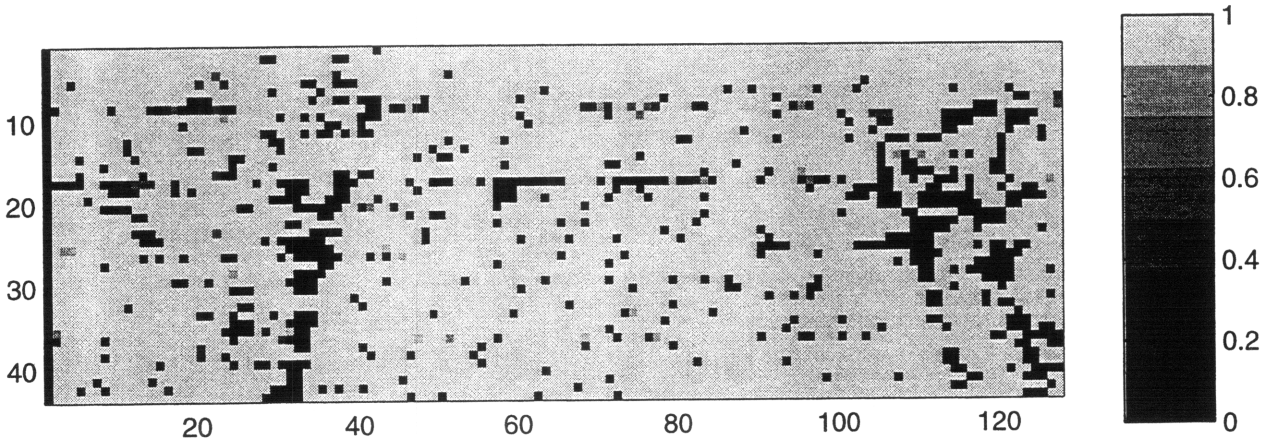

Figure 2-14: Weight image associated with the multiresolution Haar wavelet EM/ML 4×8 fit

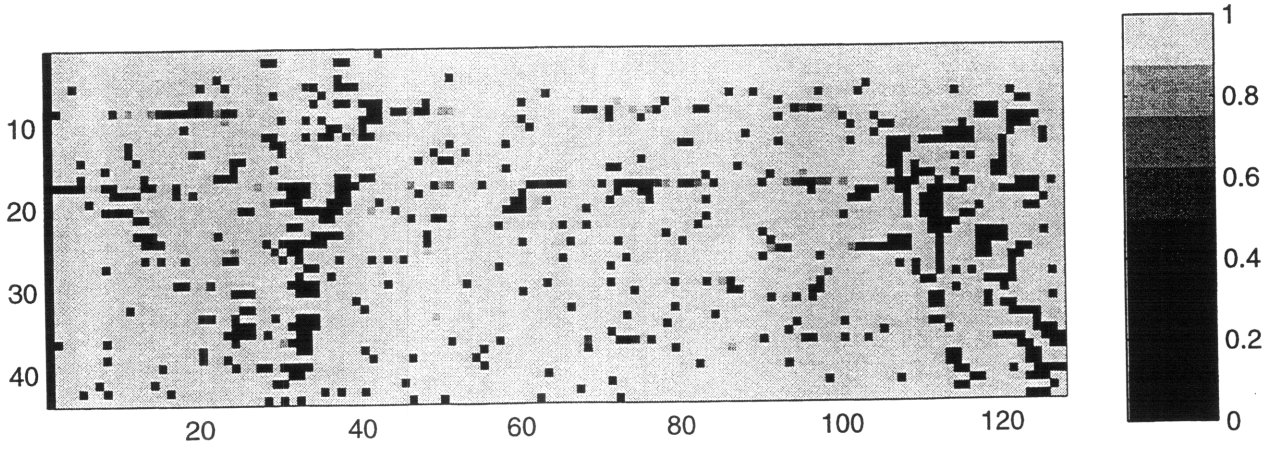

Figure 2-15: Weight image associated with the multiresolution Haar wavelet EM/ML 4×4 fit

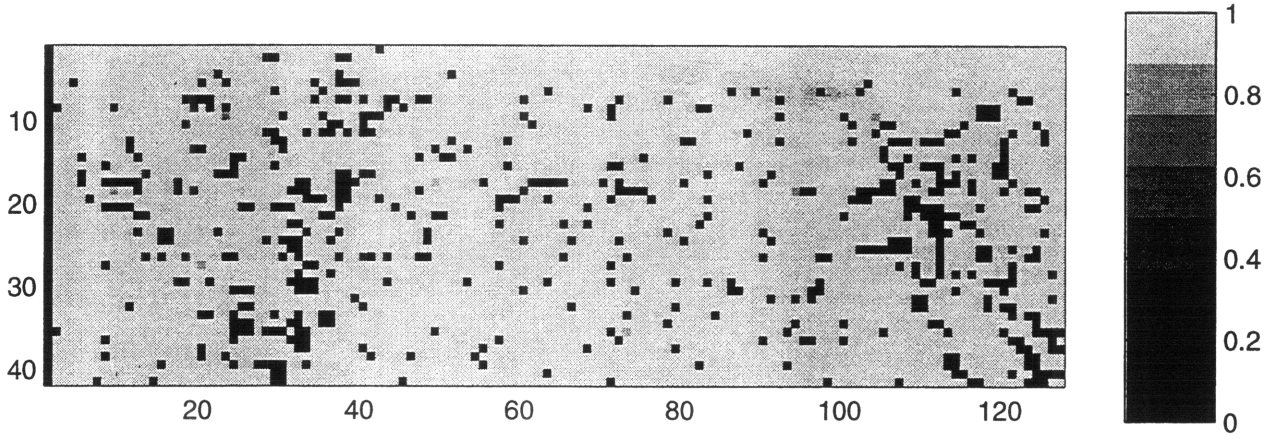

Figure 2-16: Weight image associated with the multiresolution Haar wavelet EM/ML 2×4 fit

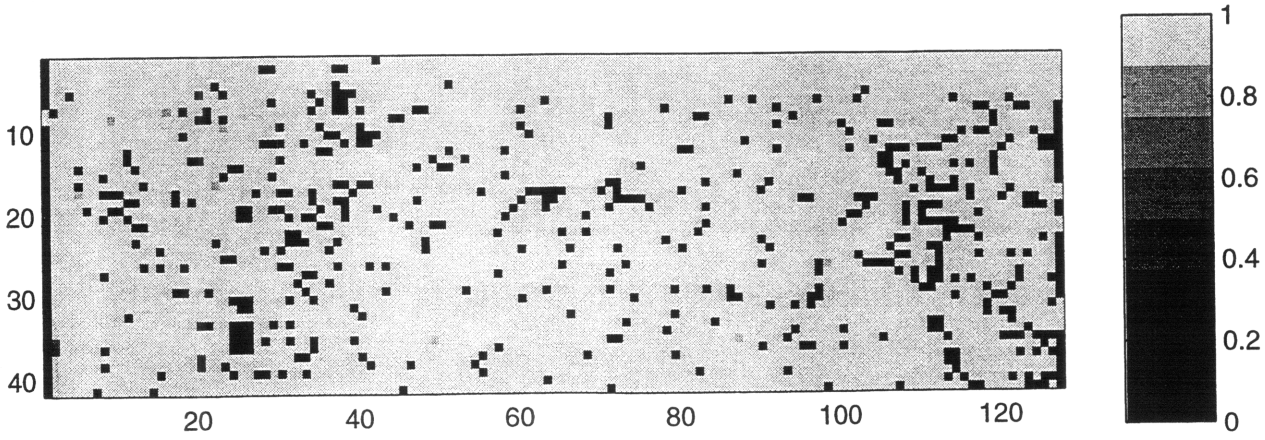

Figure 2-17: Weight image associated with the multiresolution Haar wavelet EM/ML 2×2 fit

Chapter 3

Model-Based Statistical Object Recognition System

In the computer vision literature there are many different approaches to the object recognition problem. Many of these approaches actually deal with the object verification problem; that is, they are aimed at finding if a particular object is present in an image, and if so, computing the position and orientation of the object. The position and orientation of an object in an image is usually referred to as the 'pose' of the object. A general recognition system is expected to identify and locate arbitrary objects of a model database.

In this chapter we first discuss the object recognition problem and the approaches used in this thesis to solve this problem. We then present the framework for the statistical approach, which this research rests on. The probabilistic models are presented explicitly and statistical estimation methods are developed to use these models. These methods were previously formulated by Wells [5] and used for high resolution video and synthetic range images. In this work we use these methods to establish the matching step of our object recognition system, which incorporates modifications required to process low resolution laser radar range imagery.

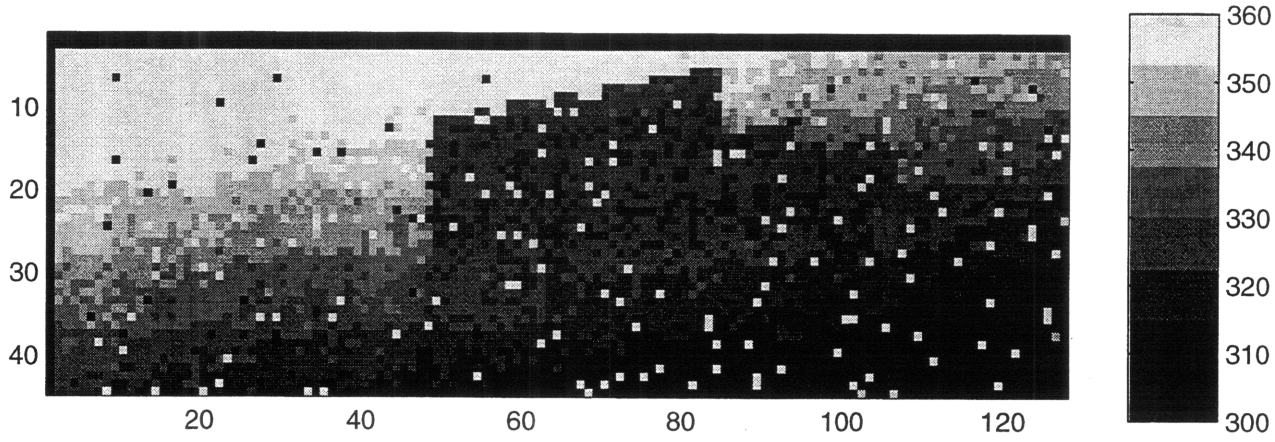

Figure 3-1: Raw range image of a truck.

3.1 Model and Feature-Based Recognition

Model-based approaches represent state-of-the art techniques for object recognition. In model-based object recognition, objects are represented by models which are known in advance and are expected to provide all the information necessary for recognition and localization. The problem then becomes using the model to locate instances of the objects in the image of interest.

Figs. 3-1 and 3-2 constitute a typical example for the model-based object recognition problem. The rendered model image in Fig. 3-2, which is a representation of the object we wish to recognize, is used to locate the object in the image in Fig. 3-1, which displays that object in a noisy degraded image together with a background.

A common approach to model-based vision is the recognition of the objects by use of localized feature matching between the model and the image. This formulation is referred to as feature-based recognition. A feature and model-based recognition approach is used in this work. Feature-based approaches, in which simple geometrical entities,

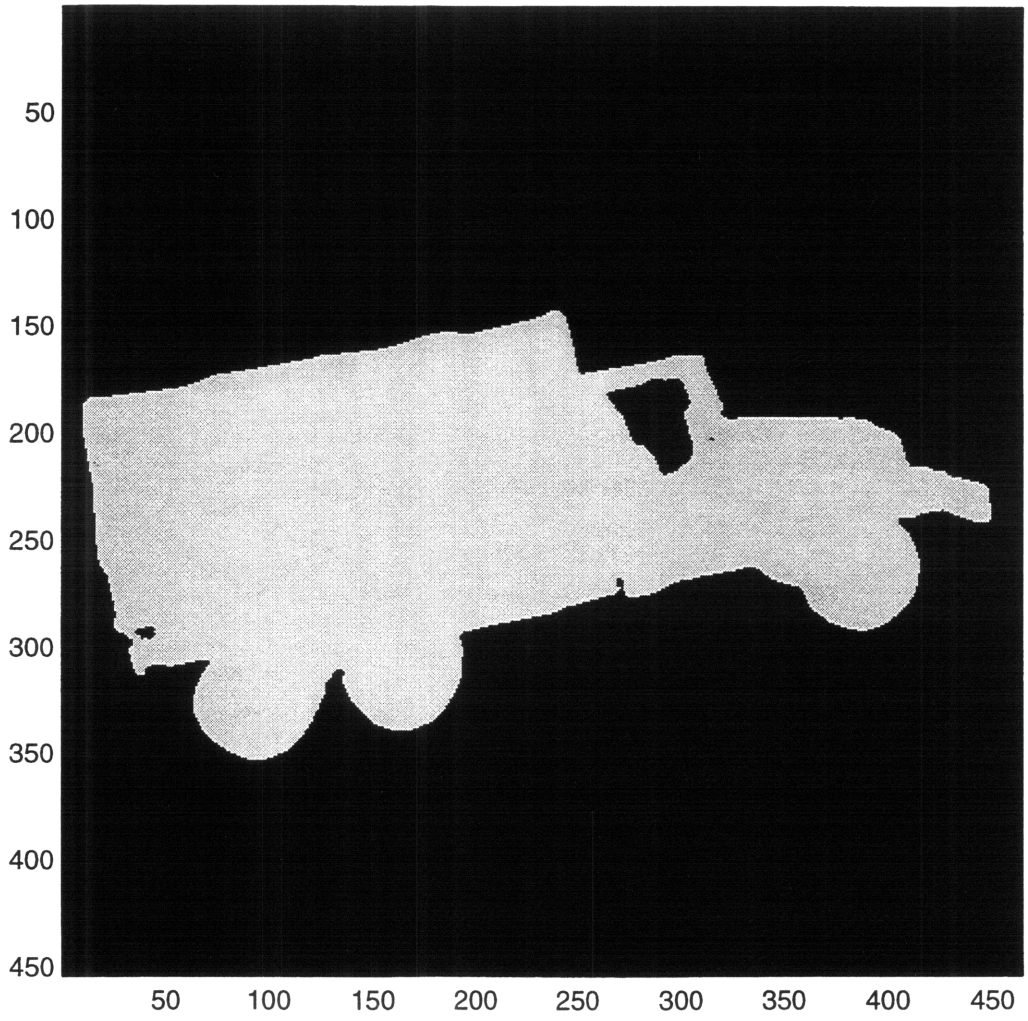

Figure 3-2: Rendered image generated from 3-D CAD model of a truck.

i.e., points, lines and curves, are used to represent the object model and the image, have been used for many years. Features are abstractions that summarize some type of structural information in an image. Different types of features can be used, such as point features, point-radius features, oriented-range features and points of maximal curvature, each conveying different levels of information. Such compact representations of the object model and the image facilitate the search algorithms involved in the recognition problem.

The main objective in feature-based recognition is to determine the optimal pairings between the model features and the image features in the sense to produce the greatest metrical consistency among the paired features. This constraint can be expressed mathematically using different measures. In this work, a statistical approach is used to develop an objective function for evaluating the hypothesized solutions to the problem.

3.2 The Statistical Approach

In this thesis, statistical methods are employed to solve the object recognition problem. Statistical models are developed to represent the uncertainty present in the problem. This approach converts the recognition problem into a well-defined optimization problem. If the domain is modeled well, the resulting statistical formulations are expected to produce reasonable results.

In order to use a statistical approach to the object recognition problem, first the components of the statistical formulation need to be identified explicitly. These models capture the essential probabilistic behavior involved in the problem and can be used to estimate pose and recognize the object. Accurate models are required to recognize the objects reliably and to interpret the results.

To be clear about the modeling procedure, it is essential to examine the various components of the formulation separately. We will start with a discussion on the types of features that can be extracted from the images. The image features are interpreted using a correspondence model. The projection model expresses the mathematics of the

deterministic transformation from the model domain to the image domain. These models are used to describe the probabilistic models of image features, which form the main component of the statistical theories of object recognition.

3.2.1 Features

As discussed in Section 3.1, feature-based object recognition is achieved by matching the extracted image and model features, which represent information about the image analyzed and the model object, respectively. Such concise representations of the salient aspects of the image and the model greatly simplify the search procedure for the optimal solution. In general, edge-based features are used since the edge contours in an image contain a great deal of information about the objects in a scene. The feature data is constructed from edge curve fragments, that is, the obtained edge curves are broken arbitrarily into fragments to form discrete features.

Different feature types can be used. In determining the feature type, there is a compromise between using complex features and detecting such features. Rich features contain more information, provide more constraints and therefore simplify recognition of objects. However, the more complex the features get, the harder it is to detect them. In the 2-D Point Feature model, the features are defined by the coordinates of the feature points extracted from the edge curve fragments. The 2-D Point-Radius Feature model is an extension that incorporates information about the normal and curvature at a point on a curve in addition to the coordinate information. The 2-D Oriented-Range Feature model was designed for use in range imagery instead of video imagery. The difference is that the inverse of the range at the discontinuity is used in this model, rather than the inverse of the curvature.

We will be working with two sources of features throughout this thesis: image features and model features. Image features are the features that we extract from the input image in which we are trying to locate the object model. Model features are the features that we extract from a model image and use to build a representation of the object model.

Background features are all image features that do not come from the object model in the image. The background features are collectively represented by the symbol, $\perp$.

In our work, we use the 2-D Point Feature model. Both the image features and the model features have information about the 2-D coordinates of the feature point extracted along the edge contours. This type of feature provides a good compromise between simplicity and accuracy.

The image to be analyzed is represented by a set of two-dimensional column vectors, denoted collectively by $\mathbf{Y}$.

$$\mathbf{Y} = (Y_1, \ldots, Y_n) \tag{3.1}$$

The object model is represented by a set of real matrices, denoted by $\mathbf{M}$,

$$\mathbf{M} = (M_1, \ldots, M_m) \tag{3.2}$$

In this formulation, image features are represented by column vectors whereas the model features are represented by matrices. This particular representation is used since it facilitates the problem formulation and solution, as explained in Section 3.2.3.

3.2.2 Correspondence Model

In any image, the features arise either from the model object we are trying to locate or from the background objects present in the scene. The statistical behavior of the features in an image depends on the source of the features. Therefore, in object recognition, it is essential to provide an interpretation of the observed image in terms of determining the source of the image features. The matching of each image feature to the model features or the background features is referred to as the **correspondence**. By interpretation of the image, we mean a set of correspondences, one for each image feature.

Correspondence

In this work, we represent the mapping from each image feature, Y_i, by the correspondence function, $\Gamma(Y_i)$, which represents the model or the background feature that corresponds to that particular Y_i. Since this mapping involves a finite number of elements, it can be represented by a finite vector, the correspondence vector,

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_n \end{bmatrix} \triangleq \begin{bmatrix} \Gamma(Y_1) \\ \vdots \\ \Gamma(Y_n) \end{bmatrix} \tag{3.3}$$

In this notation, the expression $\Gamma(Y_i)$ is equivalent to Γ_i. The correspondences are defined as a collection of variables indexed in parallel with the image features. The expression $\Gamma(Y_4) = M_5$ means that the image feature Y_4 corresponds to model feature M_5 whereas the expression $\Gamma(Y_7) = \perp$ means that the image feature Y_7 corresponds to background.

An example of a set of correspondences can be observed in Fig. 3-3. In an interpretation, each image feature is assigned to a model feature or background. However, every model feature may not be used in the correspondence set.

Probabilistic Model for Correspondences

A simple probabilistic model is used for the correspondences based on the clutter level in the image, with the intent of capturing some information bearing on correspondences before the image is compared with the object. The probability that an image feature belongs to the background may be denoted by

$$\Pr(\Gamma_i = \perp) \triangleq B \tag{3.4}$$

The remaining probability is uniformly distributed between m model features,

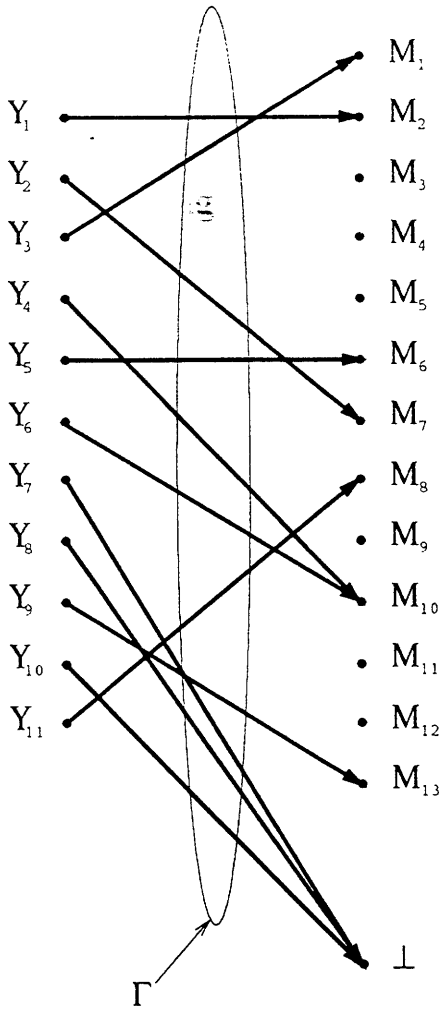

Figure 3-3: A set of correspondences between the image features and model-background features.

$$\Pr(\Gamma_i = M_j) = \frac{1-B}{m} \tag{3.5}$$

Then the marginal probability mass function of the discrete random variable Γ_i can be represented as

$$p(\Gamma_i) = \begin{cases} B & , \text{if } \Gamma_i = \perp \\ \frac{1-B}{m} & , \text{if } \Gamma_i \in \mathbf{M} \end{cases} \tag{3.6}$$

The value for B can be estimated using sample images in the domain. For instance, $B = 0.1$ would mean that 10% of the image features are expected to arise because of clutter in the image.

In this thesis, the correspondences for different image features are assumed to be independent before the image is observed. Dependent correspondence models used in a few recognition systems are discussed in [5].

By assuming independence of the components of the correspondence vector, its joint probability mass function can be expressed as the product of the marginal pmfs,

$$\begin{aligned} p(\Gamma) &= \prod_i p(\Gamma_i) \\ &= \prod_{ij:\Gamma_i = M_j} \frac{1-B}{m} \prod_{i:\Gamma_i = \perp} B \end{aligned} \tag{3.7}$$

In previous work, use of the independent correspondence model has been proven to give good performance in recognition systems. In this thesis, the independent correspondence model is used.

3.2.3 Projection Model

The positions of the image features corresponding to the target object in the image are determined completely by their correspondences and the pose when there is not any noise in the observed image. Thus in the absence of uncertainty, the projection model is a way

to express the deterministic transformation from the model features to the image features in mathematical terms.

In the most general form, the projection model can be expressed as

$$\eta_i = P(M_j, \beta) \tag{3.8}$$

where $P(.,.)$ is an arbitrary projection function, η_i represents the projection of the model feature into the image coordinate system and β represents the pose of the object.

3-D rigid body motion in space involves six degrees of freedom, three in translation and three in rotation. This six parameter pose space may be split into two parts, the first part being translations and in-plane rotation, which account for four of these parameters and the second part being the out-of-plane rotations.

In model-based object recognition systems, pose determination is a primary step to find the alignment of the object model and the image data. The pose determination problem may be greatly simplified by the use of a projection model that is linear in the pose parameters. In this way, the resulting optimization problem may be solved using the least squares approach, involving easily obtained closed form solutions.

A linear projection model can be constructed for the 2-D Point Feature model, described in Section 3.2.1. This model will essentially be two-dimensional, covering the first part of the pose space: 2-D translation, in-plane rotation and scaling in the plane. This approach will be used for recognizing 3-D objects by means of weak perspective projection, which approximates perspective projection by orthogonal projection and scaling. By this projection model, four parameters of the pose space are handled leaving only the out-of-plane rotation to be treated separately.

In the 2-D Point Feature model, the image features are represented by 2-D column vectors,

$$Y_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \text{ for } 1 \leq i \leq n \tag{3.9}$$

where x_i and y_i denote the coordinates of the image feature point. Since the pose vector involved in the projection model has four degrees of freedom, it can be represented as a column vector, $\beta \in \mathbb{R}^4$,

$$\beta = \begin{bmatrix} \mu & \nu & t_x & t_y \end{bmatrix}^T \tag{3.10}$$

Transformation of the model feature point by the pose vector is equivalent to rotation by θ, scaling by s, and translation by T, where,

$$T = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \qquad s = \sqrt{\mu^2 + \nu^2} \qquad \theta = \arctan\left(\frac{\nu}{\mu}\right) \tag{3.11}$$

This definition of the pose vector permits a linear relation to be established between the projected features and the pose vector, by representing the model features as matrices

$$M_j = \begin{bmatrix} x_j & -y_j & 1 & 0 \\ y_j & x_j & 0 & 1 \end{bmatrix}, \text{ for } 1 \leq j \leq m \tag{3.12}$$

where x_j and y_j denote the coordinates of the model feature point. The resulting linear projection model is then

$$\eta_i = P(M_j,\beta) = M_j\beta = \begin{bmatrix} x_j & -y_j & 1 & 0 \\ y_j & x_j & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \upsilon \\ t_x \\ t_y \end{bmatrix} \tag{3.13}$$

Formulating the projection model in this way, where the pose vector is a column vector and the model feature is a matrix may seem awkward at first, but it is essential to obtaining a simple solution approach for determining the pose vector.

Linear projection models can also be constructed for other feature models, such as the 2-D Point Radius Feature model, the Oriented-Range Feature model and the 3-D Linear Combination of Views features. These models are developed in [5].

3.2.4 Probabilistic Models for the Image Features

Using the framework developed so far, it is possible to construct an observation model, from which we can determine the probabilistic models for the image features to be used in the following statistical formulations. These models are presented as probability density functions for the coordinates of the image features conditioned on their correspondences and pose.

In object recognition, the features observed in a particular image correspond to either a projected model feature or a background feature with some added noise due to sensor effects, filtering and edge distortions. The probability density function (pdf) for the coordinates of the image features shows different characteristics depending on the source of these features.

Model for Background Features

Different models can be used to characterize the behavior of the image features matched to the background. Developing a satisfactory pdf for the background features is complicated, since no apriori information exists about the distribution of the background features. Therefore, it is reasonable to use a uniform density bounded by the limits of the coordinate space of the image features, which captures the maximum entropy nature of the problem. Assuming that the image features, Y_i, are 2-D vectors, the pdf for the background features is expressed as

$$p(Y_i|\Gamma_i, \beta) = \frac{1}{W_1 W_2} \text{ , if } \Gamma_i = \perp \tag{3.14}$$

in the image where W_i denotes the extent of the image coordinate system along dimension i. For instance, in an 256×256 image, the pdf becomes

$$p(Y_i|\Gamma_i, \beta) = \frac{1}{256 \times 256} \text{ , if } \Gamma_i = \perp \tag{3.15}$$

within the image.

This model represents only the expectation that the image features lie within the bounds of the image coordinate space. It is otherwise as noncommittal as possible. It has been shown that this model works well in the recognition experiments in [5]. Other models, trying to have a better understanding of the background have also been used in some works, such as in [17].

Model for Projected Model Features

For image features that are matched to the object model we are trying to locate, the observation is modeled as a deterministic projection of the model features with a specific pose determining the required coordinate transformation, and some added noise. The aggregate effect of this noise is modeled as Gaussian noise. The use of this model has been proven to be effective in [5]. Thus the projected model features are assumed to be normally distributed about their predicted position in the image and the pdf is expressed as

$$p(Y_i|\Gamma_i,\beta) = \frac{1}{2\pi \mid \psi_{ij} \mid^{\frac{1}{2}}} \exp(-\frac{1}{2}(Y_i - P(M_j,\beta))^T \psi_{ij}^{-1}(Y_i - P(M_j,\beta))), if \Gamma_i = M_j \quad (3.16)$$

where $P(M_j,\beta)$ represents the projection of the associated object feature j into the image with pose β and ψ_{ij} is the covariance matrix of the Gaussian or the normal pdf for the correspondence between the image feature i and the model feature j. The covariance matrix, ψ_{ij}, can be estimated from observations done on sample images in the domain. This procedure is discussed in more detail in Section 6.1. This pdf will be succinctly denoted

$$p(Y_i|\Gamma_i,\beta) = N(P(M_j,\beta),\psi_{ij}) \text{ , if } \Gamma_i = M_j \quad (3.17)$$

where $N(.,.)$ indicates a Gaussian or normal distribution with mean $P(M_j,\beta)$ and covariance ψ_{ij}.

Model for the Overall Observation Vector

The image features are collected in an observation vector $\mathbf{Y}$,

$$\mathbf{Y} = (Y_1, \ldots, Y_n) \tag{3.18}$$

The conditional pdfs for the image features have been specified in previous sections as

$$p(Y_i|\Gamma_i, \beta) = \begin{cases} \frac{1}{W_1 W_2} & , \text{if } \Gamma_i = \perp \\ N(P(M_j,\beta), \psi_{ij}) & , \text{if } \Gamma_i = M_j \end{cases} \tag{3.19}$$

Assuming that the image features are independent, the joint probability density of the image features can be expressed as

$$\begin{aligned} p(\mathbf{Y}|\Gamma,\beta) &= \prod_i p(Y_i|\Gamma_i,\beta) \\ &= \prod_{i:\Gamma_i=\perp} \frac{1}{W_1 W_2} \prod_{ij:\Gamma_i=M_j} N(P(M_j,\beta), \psi_{ij}) \end{aligned} \tag{3.20}$$

3.3 Alignment and Parameter Estimation

In object recognition, the primary need is to find and evaluate the alignment of the model and the image data. The alignment problem involves comparing a predicted image of an object with the actual observed image. The predicted image can be synthesized using an object model and a given pose. The parameters to be estimated in finding the correct alignment are the correspondences between the image and the model features and the pose of the object in the image. These parameters can be estimated using standard statistical methods of parameter estimation using the statistical framework developed in the previous sections.

Two different statistical formulations can be used for parameter estimation. In MAP Model Matching (MMM) method, a complete hypothesis consists of a description of the

correspondences between the image and the model features as well as the pose of the object. In contrast, the Posterior Marginal Pose Estimation (PMPE) method includes only the pose of the object, i.e., no restrictions are imposed on the correspondences between features. In this thesis, PMPE formulation of recognition is used. But, since PMPE builds on MMM, a brief presentation of both is given in the following. A thorough discussion on these methods can be found in [5],[18],[19].

3.3.1 MAP Model Matching (MMM)

In this method, maximum-aposteriori-probability (MAP) estimation is used to obtain estimates of the correspondences and pose by maximizing their posterior probability density given the observed image features,

$$\widehat{\Gamma, \beta} = \arg\max_{\Gamma,\beta} p(\Gamma, \beta | \mathbf{Y}) \tag{3.21}$$

Using Bayes' rule, the posterior probability density on these parameters given the observation of the image features is

$$p(\Gamma, \beta | \mathbf{Y}) = \frac{p(\mathbf{Y}|\Gamma, \beta) p(\Gamma, \beta)}{p(\mathbf{Y})} \tag{3.22}$$

where $p(\mathbf{Y})$, the probability of observing the image features acts only as a normalization factor because it is constant with respect to the pose and correspondence vectors.

Thus the posterior probability density can be found by using the probability density for the coordinates of the image features, conditioned on the parameters of pose and correspondence, and the prior probability density of these parameters.

The conditional probability density for the coordinates of the image features is given in Eq. 3.20. The next step is to construct a joint prior density for the parameters to be estimated. The probability model for the correspondence is given in Eq. 3.7. Prior information for the pose parameter is assumed to be given as a normal density,

$$p(\beta) = N(\beta_o, \psi_\beta) \tag{3.23}$$

where β_o is the mean and ψ_β is the covariance matrix of the normal density. In general, since there is not much information about pose prior, it is left out in formulations, resulting in a maximum-likelihood estimation for the pose.

Assuming that the correspondence and the pose are independent, the joint prior density for these parameters becomes

$$p(\Gamma, \beta) = N(\beta_o, \psi_\beta) \prod_{ij:\Gamma_i = M_j} \frac{1-B}{m} \prod_{i:\Gamma_i = \perp} B \tag{3.24}$$

Combining all this information together, we can introduce an objective function, $L(\Gamma, \beta)$, which is a scaled logarithm of the posterior probability density of correspondence and pose, $p(\Gamma, \beta|\mathbf{Y})$. The same estimates will be obtained if the maximization is carried over the objective function; i.e.,

$$\widehat{\Gamma, \beta} = \arg\max_{\Gamma,\beta} L(\Gamma, \beta) \tag{3.25}$$

where

$$L(\Gamma, \beta) \triangleq \ln\left(\frac{p(\Gamma, \beta|\mathbf{Y})}{C_1}\right) \tag{3.26}$$

for C_1 a constant.

3.3.2 Posterior Marginal Pose Estimation (PMPE)

The PMPE formulation builds on MAP Model Matching. It treats the pose parameter as the most important aspect of the problem. This is effective since it allows us to estimate the pose without directly considering the correspondences. Similar to MMM, the maximum-aposteriori estimation technique is used to find the pose estimate by maximizing the posterior probability density of the pose given the observed image features,

$$\widehat{\beta} = \arg\max_{\beta} p(\beta|\mathbf{Y}) \tag{3.27}$$

The posterior probability density of the pose may be computed from the joint posterior probability of correspondences and pose, by summing it over all correspondences, i.e., by taking the marginal over all possible matches,

$$p(\beta|\mathbf{Y}) = \sum_{\Gamma} p(\Gamma, \beta|\mathbf{Y}) \tag{3.28}$$

Using Bayes' rule, this marginal can be expressed as

$$p(\beta|\mathbf{Y}) = \sum_{\Gamma} \frac{p(\mathbf{Y}|\Gamma, \beta)p(\Gamma, \beta)}{p(\mathbf{Y})} \tag{3.29}$$

This expression takes the following form under the feature independence and correspondence-pose independence assumption,

$$p(\beta|\mathbf{Y}) = \frac{1}{p(\mathbf{Y})} \sum_{\Gamma_1} \cdots \sum_{\Gamma_n} \prod_i p(Y_i|\Gamma, \beta) \prod_i p(\Gamma_i)p(\beta) \tag{3.30}$$

or equivalently,

$$p(\beta|\mathbf{Y}) = \frac{p(\beta)}{p(\mathbf{Y})} \sum_{\Gamma_1} \cdots \sum_{\Gamma_n} \prod_i [p(Y_i|\Gamma, \beta)p(\Gamma_i)] \tag{3.31}$$

Simplifying this expression by breaking factors out of the product, one at a time, yields

$$p(\beta|\mathbf{Y}) = \frac{p(\beta)}{p(\mathbf{Y})} \prod_i p(Y_i|\beta) \tag{3.32}$$

where

$$p(Y_i|\beta) = \sum_{\Gamma_i} [p(Y_i|\Gamma_i, \beta)p(\Gamma_i)]$$

$$= p(Y_i|\Gamma_i = \perp, \beta)\Pr(\Gamma_i = \perp) + \sum_j p(Y_i|\Gamma_i = M_j, \beta)\Pr(\Gamma_i = M_j) \quad (3.33)$$

Substituting the probability density functions developed in the previous sections leads to

$$p(Y_i|\beta) = \frac{1}{W_1 W_2} B + \sum_j N(P(M_j, \beta), \psi_{ij}) \frac{1 - B}{m} \quad (3.34)$$

Similar to the previous section, an objective function for PMPE, $L(\beta)$, may be defined as the scaled logarithm of the posterior probability density of the pose, $p(\beta|\mathbf{Y})$,

$$L(\beta) = \ln\left(\frac{p(\beta|\mathbf{Y})}{C_2}\right) \quad (3.35)$$

where C_2 is a constant. After several manipulations, and using the linear projection model, this objective function can be expressed as

$$L(\beta) = -\frac{1}{2}(\beta - \beta_o)^T \psi_\beta^{-1}(\beta - \beta_o) + \sum_i \ln\left[1 + \sum_j \frac{W_1 W_2}{m} \frac{1 - B}{B} N((M_j\beta), \psi_{ij})\right] \quad (3.36)$$

Close examination of this objective function reveals that it measures the degree of alignment between the image and the model data, penalizing the deviations from the predicted location in the image.

The most important property of this formulation is that the resulting objective function for evaluating a pose hypothesis is a smooth function of the pose. As explained in more detail in the next section, the expectation-maximization (EM) algorithm may be used to search for local maxima. Therefore, computation involved in PMPE formulation of object recognition shares the same mechanism with that of ML range imaging. Both methods use the EM algorithm to find statistically optimal estimates of the variables of interest. The zero gradient condition results in a nonlinear equation which can be solved iteratively in successive expectation and maximization steps. If the initial pose estimate is sufficiently good, i.e., if it is on the likelihood "hill" associated with the global

maximum, then the EM algorithm will converge to that global maximum.

3.3.3 Expectation-Maximization Algorithm

The expectation-maximization (EM) algorithm has been presented in detail in Section 2.2.1. In this section, a specific formulation of the EM algorithm is presented to be used for PMPE estimation. In PMPE, the pose is estimated by maximizing the posterior probability density given the image,

$$\widehat{\beta} = \arg\max_{\beta} p(\beta|\mathbf{Y}) \tag{3.37}$$

This estimate should satisfy

$$\frac{\partial}{\partial\beta} \ln(p(\beta|\mathbf{Y}))|_{\beta=\widehat{\beta}} = 0 \tag{3.38}$$

which is the necessary condition for an extremum at $\beta = \widehat{\beta}$. The posterior probability density for the pose given the image is given in Eq. 3.32. Computing the gradient and setting it equal to 0 leads to

$$\psi_{\beta}^{-1}(\widehat{\beta} - \beta_o) + \sum_i \frac{\frac{1-B}{m}\sum_j N((M_j\widehat{\beta}), \psi_{ij}) M_j^T \psi_{ij}^{-1}\left(Y_i - M_j\widehat{\beta}\right)}{\frac{B}{W_1W_2} + \frac{1-B}{m}\sum_k N((M_k\widehat{\beta}), \psi_{ik})} = \mathbf{0} \tag{3.39}$$

which can be expressed compactly as

$$\psi_{\beta}^{-1}(\widehat{\beta} - \beta_o) + \sum_i \sum_j w_{ij} M_j^T \psi_{ij}^{-1}\left(Y_i - M_j\widehat{\beta}\right) = \mathbf{0} \tag{3.40}$$

where

$$w_{ij} = \frac{N((M_j\widehat{\beta}), \psi_{ij})}{\frac{B}{1-B}\frac{m}{W_1W_2} + \sum_k N((M_k\widehat{\beta}), \psi_{ik})} \text{ for all } i, j \tag{3.41}$$

Note that w_{ij} is the conditional probability that image feature Y_i and model feature M_j correspond, given that the most recent pose estimate is correct.

Eq. 3.40 looks like a linear equation in $\widehat{\beta}$, but it is not since the weights, w_{ij}, are functions of $\widehat{\beta}$. Therefore, to solve for $\widehat{\beta}$, the EM algorithm is used, which iterates between the following two steps producing a sequence of estimates for which the associated objective function values are monotonically increasing:

1. The weights are computed using the most recent pose estimate by Eq. 3.41. This is called the **expectation step**. It corresponds to computing the probability that the ith image feature and jth model feature correspond. At the end, the weights provide continuous-valued estimates of the correspondences given the image.

2. Eq. 3.40 is solved as a linear equation for a new pose estimate $\widehat{\beta}$ assuming the current estimates of the weights, w_{ij}, are correct. This is called the **maximization step** in the original formulation of the EM algorithm [16] since it corresponds to computing the maximum-likelihood estimate.

At the end of EM iterations, good measures of feature correspondences are obtained in the expectation step, although they have been left out in the original formulation of PMPE.

The EM algorithm can be started in either step, depending on whether an initial set of weights or an initial pose is available. In this work, we assume that a good initial pose is provided and we are interested in refining the value of the pose vector by a local search in the pose space.

Chapter 4

Object Recognition System Characteristics

The objective in this thesis is to develop a target recognition system capable of recognizing military vehicles in registered noisy range images produced by airborne laser radars. This system is intended to contain modules having quantitative performance criteria for optimum use of sensor data. Statistical detection and estimation theory provides rigorous approaches to develop these models so that overall system optimization may be possible.

Toward that end, the previously developed fast EM/ML algorithm could serve as an essentially optimal preprocessor. The resulting range profiles would be used as the input to the successive modules in the system. For the object recognition module, we have chosen to employ a model-based statistical method, in particular the PMPE object recognition algorithm, used previously on a totally different domain including video images with high resolution and appreciable amount of clutter in the background. Figs. 4-1 and 4-2 demonstrate two examples to compare the type of imagery for which the algorithm has been successfully used before with the type we intend to work on.

Our research will combine these modules to develop a system suitable for real laser radar range images by doing the required modifications and by constructing the intermediate steps. The theoretical framework required for these modules has been presented

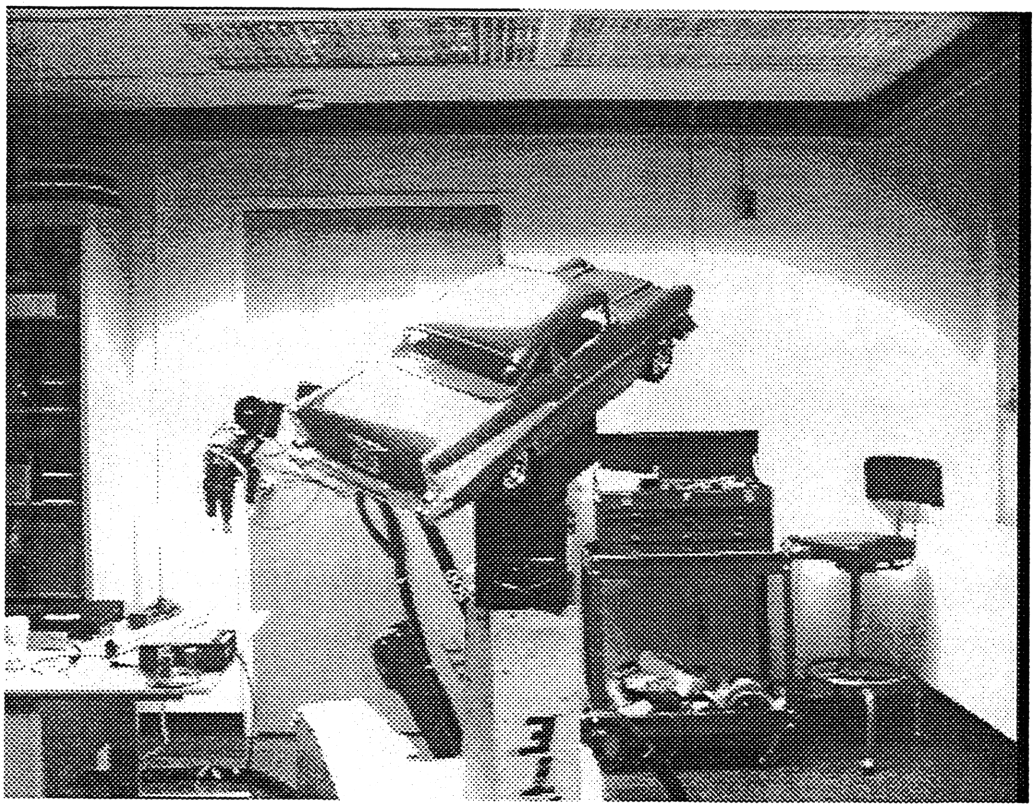

Figure 4-1: Video image.

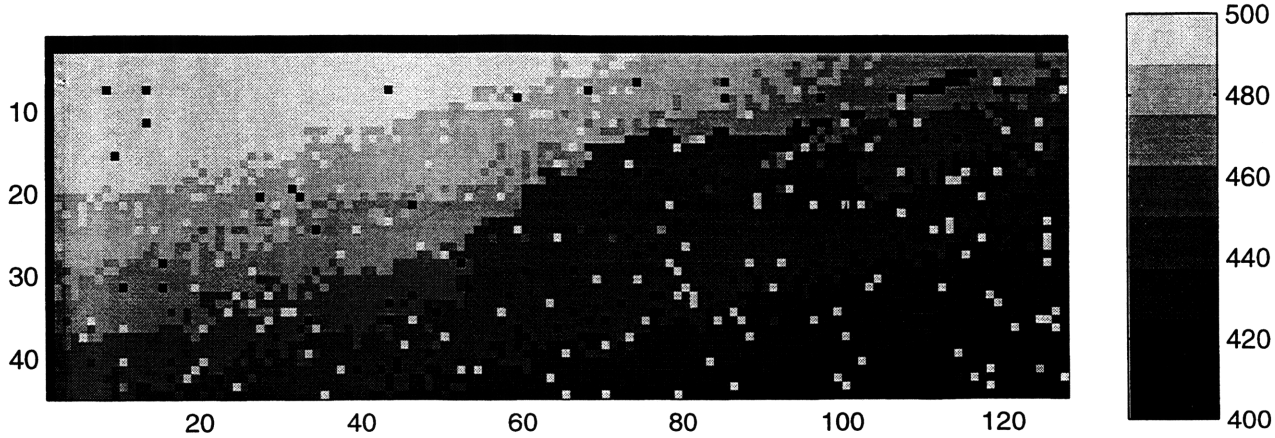

Figure 4-2: Raw range image.

in Chapters 2 and 3. In this chapter an overview of the overall object recognition system will be given. The characteristics and the parameters of the inputs to the system, namely the range images and the model, will be presented. In the latter part of this chapter implementation details suitable to achieve alignment of the object models with the imagery of interest will be discussed.

4.1 Overview of the Object Recognition System.

The object recognition system contains modules that are applied sequentially to raw sensor data. The key components of the overall object recognition system are summarized in Fig. 4-3. The algorithm works in a parallel fashion, assigning two independent processors dedicated to processing the two inputs of the system, the range image and the object model. The output of the system is the value of the objective function of the PMPE matcher, which gives an indication of the degree of the alignment between the image and the particular object model. For each image of interest, the models in the 3-D Model Library, that account for the possible military targets in the image, are applied to the system. The corresponding scores of alignment at the output of the system are compared to find the type of the located target in the image.

The noisy, low resolution laser radar range image is first processed by the fast EM/ML algorithm to extract the finest scale information adequately supported by the raw observation data, while simultaneously suppressing the range anomalies. The resulting range profile is segmented to isolate the target to be identified from the background as accurately as possible. The purpose of the following feature extraction module is to distil the essential characteristics needed to identify the target in the segmented raw images. In order to accurately and concisely model arbitrarily shaped objects, the images are represented by features extracted from the edge contours in the image. This provides a transition from the raw sensor data to a different domain involving a set of relevant edge-based features. Then the range image is submitted to the recognition stage to de-

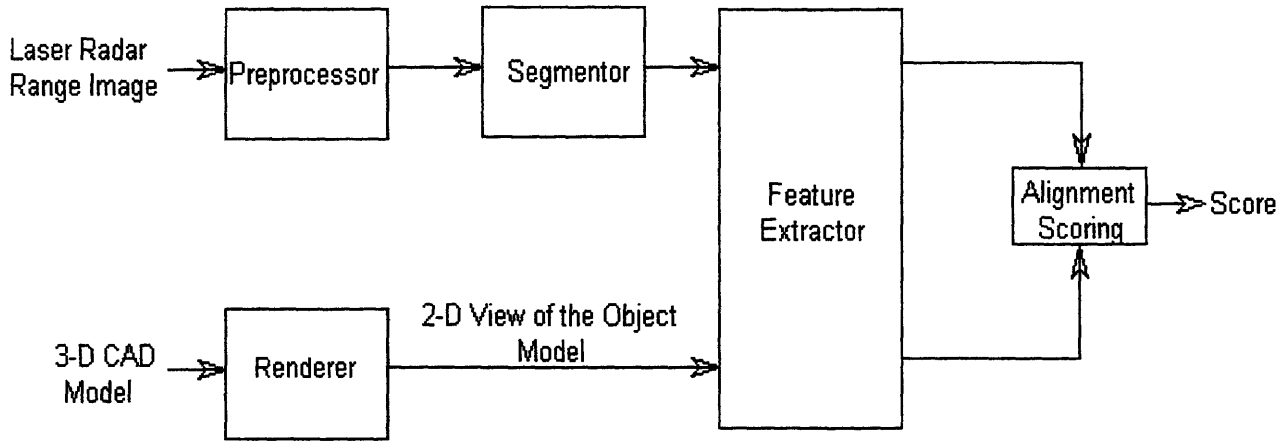

Figure 4-3: Block diagram of the overall object recognition system.

termine whether or not the target in the image corresponds to an object that is known to the object recognition system.

The 3-D CAD models that represent the separate objects that may be present in the image are first processed, one at a time, by the available renderer program to synthesize binary, 2-D views of the object at an arbitrary pose. These 2-D images have much higher resolution than the range images used. Therefore, the resolution of these images are reduced by grouping pixels and assigning to each cluster a value consistent with the majority of the pixels. The same feature extraction module used for the range images is applied to the resolution-degraded 2-D model image to construct a similar compact representation for the object model.

The matching of the object model to the image is done by the PMPE recognition module. Essentially, the object model is projected into the image plane with the predicted pose and then compared with the actual image. The output of this algorithm is the value of the objective function of the PMPE module, which provides a level of alignment of the object model with the range image.

4.2 Laser Radar Range Imagery

The laser radar imagery used in this work has been collected by MIT Lincoln Laboratory, as part of the Infrared Airborne Radar (IRAR) program and is available via the IRAR data release [20]. All of the data sets were collected from experimental ground-imaging sensors aboard a Gulfstream G-1 aircraft. The sensors are divided into two independent units according to the area scanned as the aircraft moves along its flight trajectory on a data collection measurement: a forward-looking unit with the sensor field-of-regard pointing ahead and somewhat below the direction of flight and a down-looking unit with the sensor field-of-regard pointing directly at the ground below the aircraft.

The forward-looking optical sensor suite has two operating modes: linescan mode and framing mode. In linescan mode, the scanning mirror scans in azimuth with the 12-element detector array providing an image of 12×3840 pixels. The framing mode provides more rapid imaging of a smaller area than does the linescan mode, providing images with visually apparent specific objects. The down-looking system which supports a single scanning mode, operates similar to the linescan mode in the forward-looking unit.

More specifically, the imagery particularly used in this thesis was produced by a radar system carried on the aircraft equipped with laser intensity and range sampling, Doppler sampling and video recording, operating in framing mode. As the aircraft flew towards the target, large sets of measurements have been taken, each set composed of a video, range, intensity and passive IR images in order.

In this research, we are mainly interested in processing forward-looking framing-mode range images. Fig. 4-4 illustrates a sample video image containing an army tank viewed from the back. The corresponding range image is shown in Fig. 4-5 and illustrates a tank situated in sloping, but otherwise featureless terrain. It is apparent from these images that the field of view of the laser radar range sampler is much more focused than the field of view of the video recorder.

The range data is a 45×128 pixel image. The gray shade of each pixel represents the

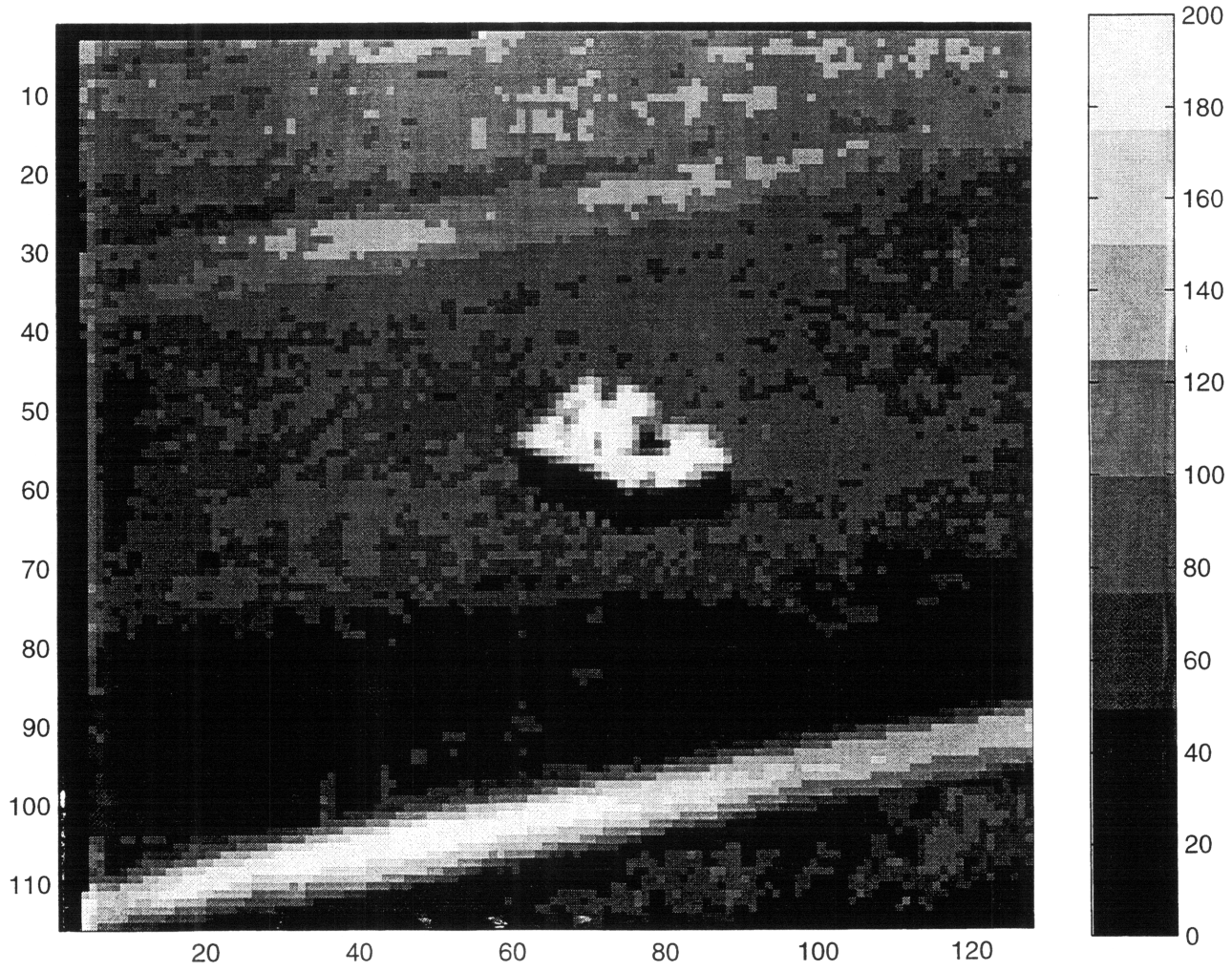

Figure 4-4: Video image of a tank viewed from the back.

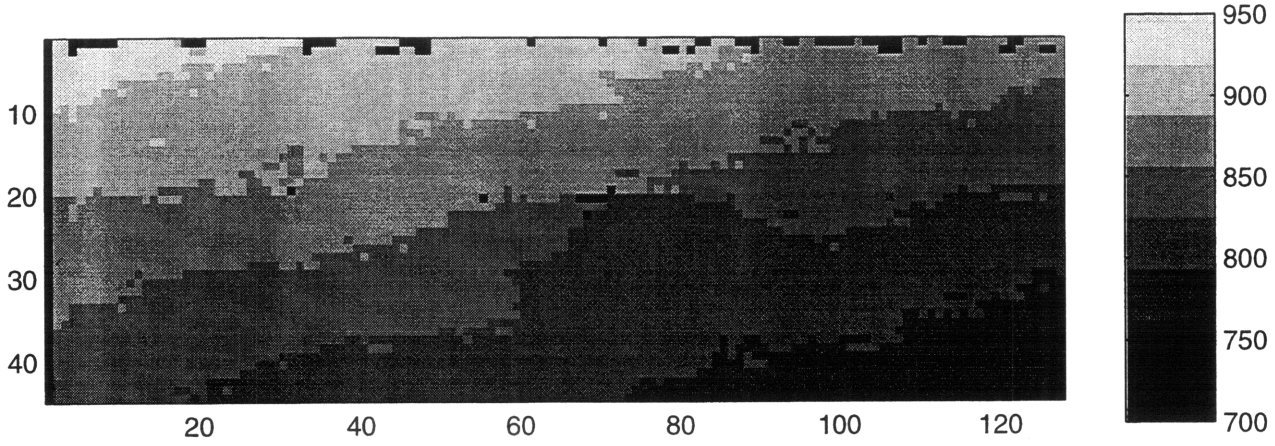

Figure 4-5: Range image of a tank viewed from the back.

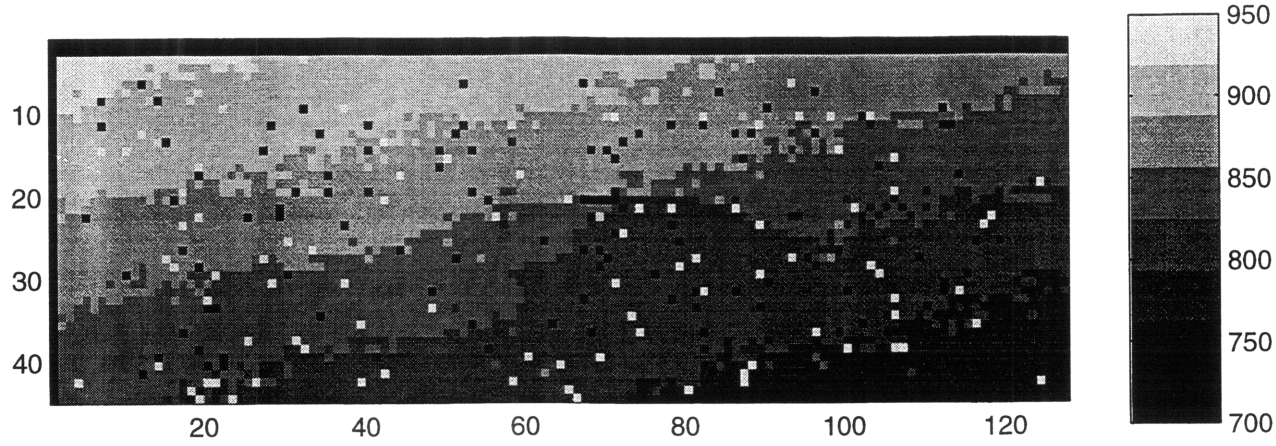

Figure 4-6: Range data of a tank, artificially created from the range truth by addition of statistically independent, zero mean Gaussian noise to each pixel and random creation of anomalies.

distance in range bins measured by the laser radar. The calibration bar in Fig. 4-5 shows that the range image is between 700 and 950 range bins away from the laser radar.

According to the single pixel statistical model used in this research, the range image involves some anomalous pixels and a certain amount of Gaussian noise. However, this image was taken under high CNR conditions and therefore the fraction of anomalous pixels seems to be small; it is almost anomaly free. Therefore, this image is taken to be the range truth from which realistic, simulated raw images can be produced. Actually, this is not the case since the Gaussian noise due to the local oscillator shot noise is always present in the range measurements and Fig. 4-5 is in fact, $R^* + n$, where n is a Q-D column vector composed of independent Gaussian random variables, each with zero mean and δR^2 variance. However, this range image is assumed to give the true range values throughout this work.

From this range truth, R^*, range data, R, is produced in accordance with the statistical model presented in Section 2.1. In particular, for each raw image generated from this range truth, local Gaussian noise with standard deviation, $\delta R = 2$, is added to each pixel. Then the anomalies are simulated with a 5% anomaly rate using a Bernoulli process to select the pixels to be made anomalous. For each selected pixel, a random range value is chosen uniformly across the range uncertainty interval given by the range gate width, $\Delta R = 1524$ range bins. The result is a raw laser radar range image conforming to the developed statistical model with a known range truth. The resulting range image, R, is shown in Fig. 4-6. Note that in the image in Fig. 4-6 the top and left edges are shown as solid black lines, corresponding to pixels where the laser radar had recorded 'no reading'. These pixels were set to zero producing these edge effects in the image.

4.3 3-D CAD Models

Representation of the objects is an important issue in model-based object recognition. For recognition, it is essential to predict the image features that will appear in an image of

the object using the object model itself. Usually, 3-D data structures that may be derived using CAD programs are used for 3-D recognition. This is due to the fact that computer graphics techniques can be used to synthesize reasonable images from 3-D models in any pose desired.

Such 3-D representations are useful especially for polygonal objects since it is easy to determine how the object will appear at different poses. On the other hand, for objects that are smoothly curved, it is difficult to predict how the object will look like in an image at a particular pose using a 3-D representation. Another alternative for modeling the objects is to use an image-based approach, in which the images of the object are used to construct the model in a way that covers the space of poses that the object may assume. Ullman and Basri [25] presented such an approach, known as linear combination of views to construct a model of an object using a number of 2-D views interpolated to synthesize images at a given pose. An image-based approach has been used in the previous work in [5], assuming the interpolated view has already been generated.

In this thesis, we have chosen to use 3-D CAD models to represent the targets in the images since an image based approach is not as effective for this domain as it is for the video image domain. The drawback of using 3-D CAD models is not being able to select the low resolution edge features that are likely to appear in an image as opposed to a view-based approach. However, for the targets of interest, alignment experiments have shown that the 3-D CAD models provide satisfactory representations for the objects.

The required 3-D models for the specific military vehicles, that might be present in the area scanned by the laser radar, have been found by Web search and have been purchased from the REM 3-D Model Bank in the format appropriate for use on existing rendering programs. Each model was available at three different levels of detail, we have chosen to use moderate level of resolution for our purposes.

It is possible to render the object models with an arbitrary pose using the "Inventor[1]" rendering program. In this way, 2-D images of the object from any viewing angles can

[1] "Inventor", Silicon Graphics Inc., Mountain View, CA.

be generated. Figs. 4-7 and 4-8 illustrate binary images corresponding to two different renderings of the same object model. Other synthesized 2-D images for two other models in our model library can be seen in Figs. 4-9 and 4-10.

4.4 Alignment of Image and Model Features

In object recognition, the main task is to find and evaluate the alignment of the image and the model data. The general problem of alignment involves comparing a predicted image of the object model with the actual image. Given an object model and a particular pose, the resulting image can be predicted. The predicted image can then be compared to the actual image directly and if the object model and the pose are correctly estimated, the predicted and actual image should match.

For our case, the 2-D image generated from the object model and the raw range data are formed by totally different mechanisms. Basically, they are representations of the target in different domains. Figs. 4-11 and 4-12 show the noisy raw range image of an M60 tank on a sloping background and the image generated from the corresponding object model respectively. Note that the pose with which the object model is rendered in Fig. 4-12 is not the actual pose of the object in the range image in Fig. 4-11.

Hence, the procedure described for the image-based approach in Chapter 3 needs to be modified for the alignment of the features extracted from the laser radar range image and the rendered object model.

4.4.1 A New Coordinate system for the Features

The features of both the analyzed range image and the rendered object model are extracted from the edge contours. In the 2-D Point feature model, as explained in Section 3.2.1, each feature contains information about the x and y coordinates of the extracted feature point, using number of pixels as a measure along each dimension. However, since the two images are obtained using different imaging processes, it is reasonable to use

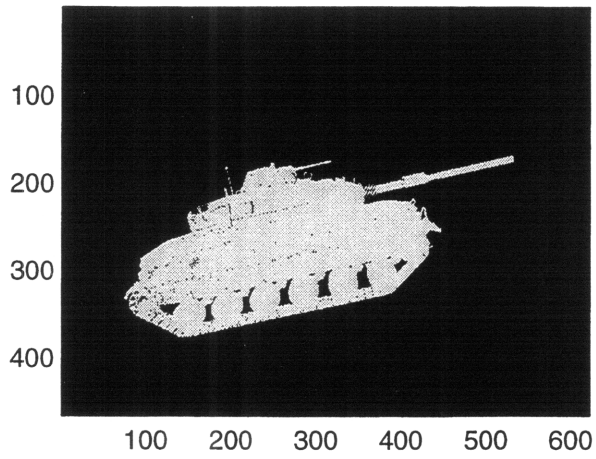

Figure 4-7: Rendered image generated from the 3-D CAD model of an M60 tank.

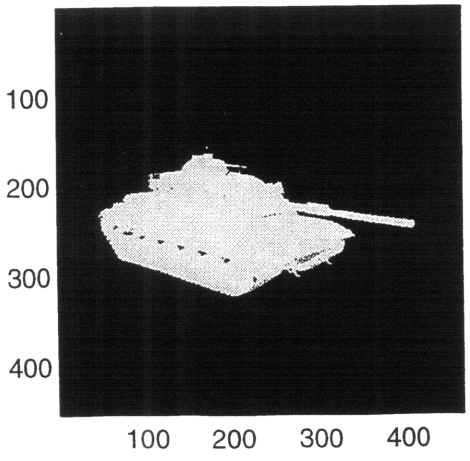

Figure 4-8: Rendered image generated from the 3-D CAD model of an M60 tank.

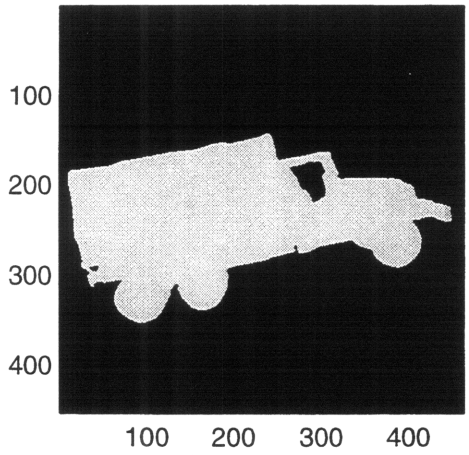

Figure 4-9: Rendered image generated from the 3-D CAD model of a truck.

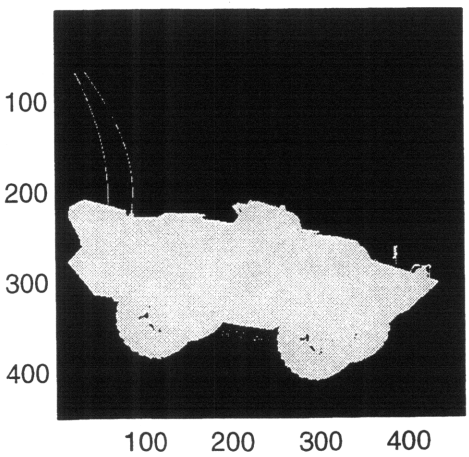

Figure 4-10: Rendered image generated from the 3-D CAD model of an armored personnel carrier.

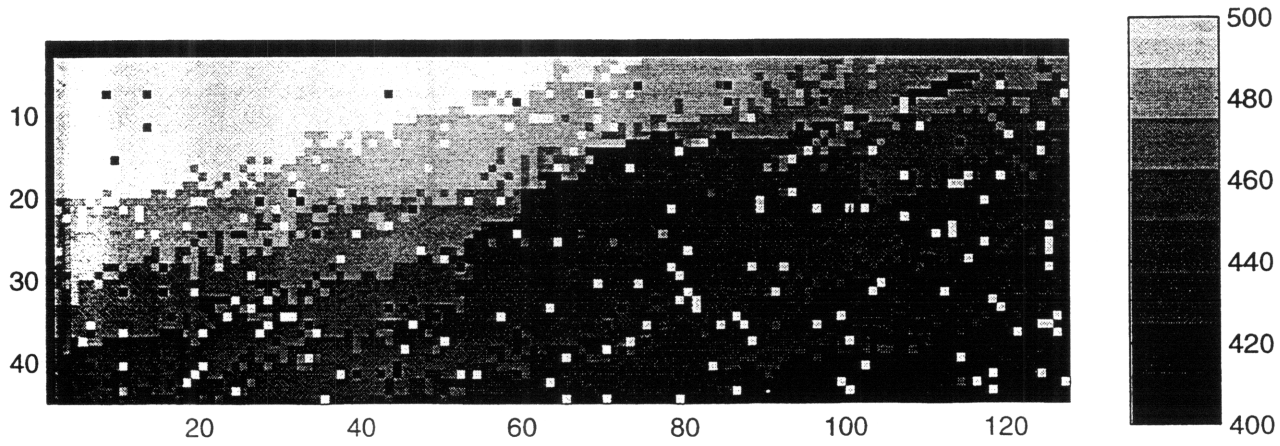

Figure 4-11: Raw range image of an M60 tank situated on a sloping background.

information about the actual dimensions (in meters) instead of using the number of pixels as the measurement unit in the matching process. Therefore in order to make a fair comparison between the features in the range image and the object model, both must be mapped to a 3-D coordinate system in which they represent the object with its actual size. This can be done by figuring out what the pixel sizes are in meters for each image.

The features of the range image are located in 3-D rectangular coordinate system using the various settings used by the laser radar to produce the imagery. The distance in the range image is represented by range bins. The size of a range bin is 1.1 meters, the range gate offset is 1400 feet=427 meters, and the range gate width, ΔR , is 5500 feet=1524 range bins=1676 meters. The origin is placed at the center of the image and the frame of reference for the coordinate system is set as shown in Fig. 4-13. The distance corresponding to the z coordinates of locations of each pixel is calculated via

$$z = (\#\text{ of rangebins}) \times (1.1\text{m}) + \text{Range Gate Offset(in m)} \tag{4.1}$$

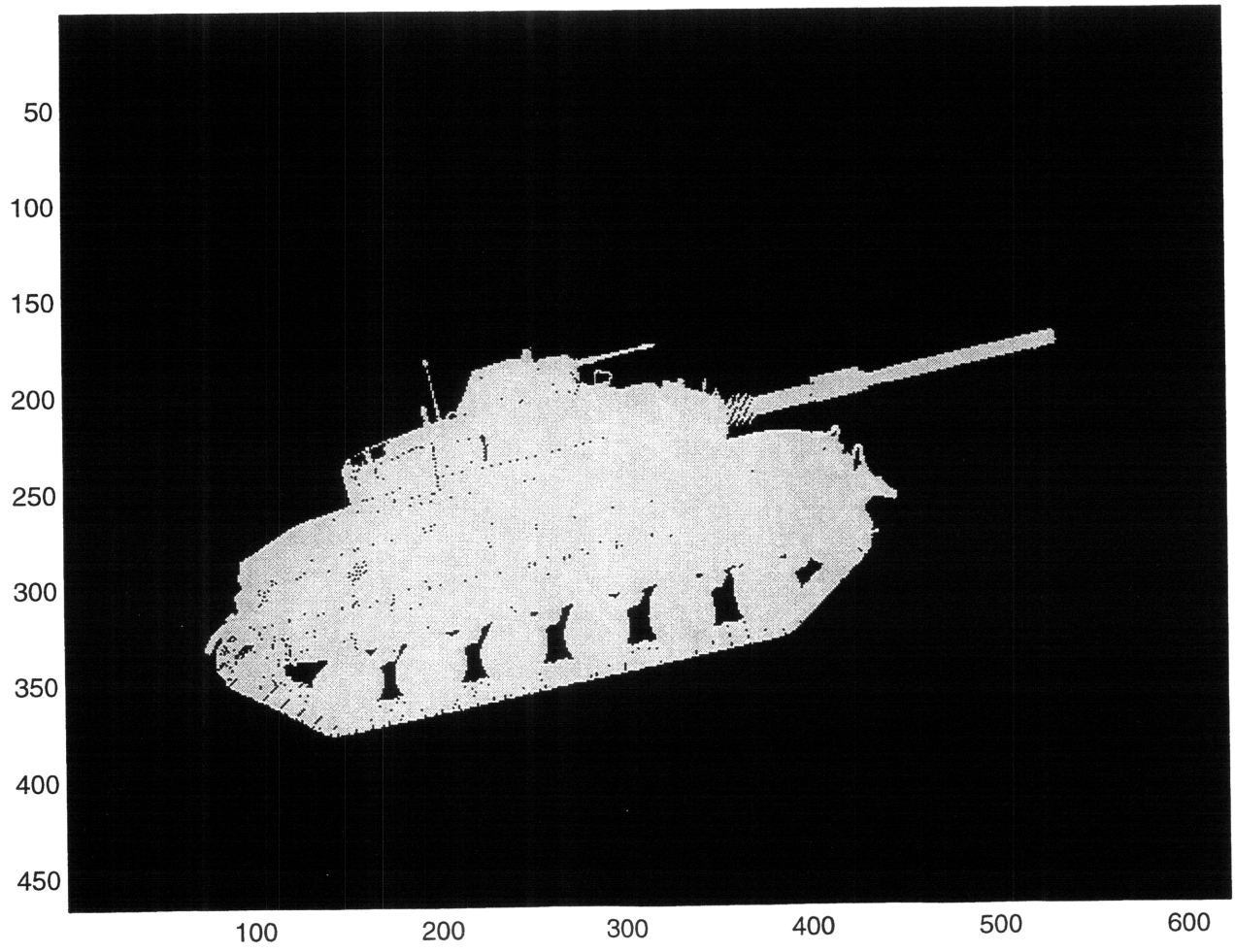

Figure 4-12: Rendered image generated from the 3-D CAD model of an M60 tank.

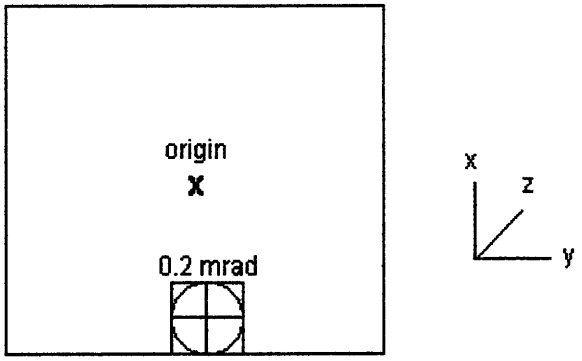

Figure 4-13: Sensor coordinate system.

The laser radar has a 0.2-mrad-full-angle instantaneous field-of-view per pixel. This determines the size of each pixel at a certain range value, from which the x and y coordinates of the pixels, corresponding to the image features, can be deduced, i.e., a pixel at 1 km is 20 cm in length and width.

Similarly, the 3-D coordinates of the model features in the rectangular coordinate system may be found from the parameters used by the available renderer program in synthesizing the 2-D views of the object. However, the renderer used in this thesis does not reveal quantitative information about the position of the object with respect to the camera or the actual dimensions of the object, from which the pixel size in meters can be deduced. Instead, information about the actual size of the targets in meters have been supplied by the REM 3-D Model Bank, where the corresponding 3-D models were purchased. This information has been used to determine the pixel size for the rendered images. To be more accurate, the width or height of the target in terms of number of pixels is compared with the actual dimensions in meters to find the pixel size. This may give incorrect results if the rendering of the target object involves out-of-plane rotations. Therefore, the pixel size is calculated for an initial pose involving no out-of-plane rotations, after which the required out of plane rotations are applied keeping the other four parameters of the pose constant. The origin is assumed to be at the center of

the image and the same frame of reference used for the range image coordinate system is used for the rendered range image. Although not perfectly precise, this approach yields a reasonable approximation for the location of the x and y coordinates of the model features in the 3-D coordinate system. Note that the z coordinates of the model features cannot be deduced from this information. This is not a problem since only the in-plane pose parameters are explicitly modeled in our object recognition system. Hence, the z coordinate information is not used in the alignment process.

4.4.2 Transformations in Pose Space

In this research, we are trying to develop an object recognition system intended to handle all six parameters of the rigid body motion. A linear projection model incorporating translation in three dimensions and in-plane rotation has been defined in Eq. 3.13 for the model features. However, for the out-of-plane rotations, a projection model linear in the parameters of the transformation cannot be defined. Therefore, an objective function in which the out-of-plane rotation pose parameters, denoted by θ_x and θ_y, are incorporated cannot be optimized efficiently. The subscripts in the notation illustrate the axis of rotation consistent with the coordinate system defined in Fig. 4-13. Instead, these parameters are determined by forming a set of discrete hypotheses in the form of 2-D synthetic views of the 3-D model. In-plane translation, rotation and scaling of the views are used to approximate full three-dimensional motion of the object.

We used the rendering program to vary the out-of-plane rotation parameters and create 2-D synthetic images from the 3-D model. These "standard orientations" cover the space of 3-D rotations whose axis is perpendicular to z axis. A certain number of views, $1 \leq k \leq K$, are catalogued and the same algorithm is applied to each one of them to obtain the view for which the algorithm gives the highest score, which is determined by the value of the objective function. This view determines the out-of-plane pose parameters of the object.

The remaining four parameters can be handled by defining an explicit projection

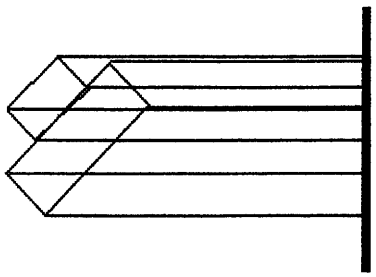

Figure 4-14: Orthographic (parallel) projection of features onto a reference plane.

model, an example of which is the linear projection model defined for the 2-D Point Feature model in Section 3.10. These models are essentially 2-D since the transformations comprise translation, rotation and scaling in the plane within weak perspective projection which approximates perspective projection by orthogonal projection with scaling. Therefore, in the search process, the image and the model features should contain information about only x and y coordinates. This is equivalent to applying orthographic (parallel) projection to features located in a 3-D coordinate system, by which we obtain the actual size of the image projected onto a reference plane, as shown in Figure 4-14.

Use of range data as the input to our recognition system results in further simplification in the search process. The pose vector, which we are trying to estimate by a local search in pose space, has been defined to be a vector having four degrees of freedom in Eq. 3.10. The associated linear projection model represents transformations in the form of in-plane translations, in-plane rotation and scaling which represents information about translation in z direction. Since in the case of range data we have the range information available, we can get rid of the scaling parameter search in our algorithms. This reduces the dimension of the search space from four to three; the only parameters to be estimated are the in-plane translations and the in-plane rotation. A projection model linear in these three parameters will be presented in the next section. Explicitly modeling translation and rotation in the plane combined with considering the appearance of an object from the possible viewing directions and with the available range information approximates

the full, six-dimensional transformation space.

Using discrete hypotheses for out-of-plane pose parameters and discharging z-translation result in a reduction of the search space. The idea is similar to that used in [24], in which "marginalization" over one of the dimensions is proven to be a powerful tool for reducing the complexity of the optimization problem. In that work, search in 2-D space results in starting values to be used in the subsequent 3-D search. In our case, however, an additional search is not required because the range information is already available in the images. At the end of PMPE optimization by which the optimum values for the in-plane translation and rotation parameters are found the z-translation parameter value is recovered using the range information available, specifically by the average of the range values of the pixels corresponding to the target.

4.4.3 A New Projection Model

Use of range images as the input to the object recognition system allows us to do search in a pose space having only three dimensions. Neglecting the scaling parameter in the search and dealing only with the three in-plane parameters is probably one of the most important characteristics of the developed object recognition system in terms of classification between distinct object models. Models of considerably different size with respect to the actual target in the image are constrained to have poor degree of alignment since they are not allowed to be scaled in the course of the search process. If scaling were allowed during alignment, the matching algorithm would be allowed to align the main boundaries of the target in the image and the model in process, no matter how disparate their true sizes were. This would drastically increase the probability of misclassification for low resolution data. For instance, two different tanks of different sizes can be confused in the recognition algorithm, since when scaled, the overall shape of the target in the image generated by the features may be insufficient to distinguish between the two objects. Excluding the scaling parameter in the process precludes the possibility of confusing targets which, although similarly shaped, have different dimensions in the real

world.

Nevertheless, this improvement leads to another discrepancy in the system. The linear optimization with respect to the pose vector of the PMPE objective function via the EM algorithm is one of the most attractive features of the PMPE algorithm. For the search in 3-D pose space, it is impossible to define a projection model which is linear in the three in-plane parameters of the pose preventing us from solving the problem via linear optimization.

To overcome this difficulty, we have employed the following approximation. Note that the algorithm is provided with a sufficiently good initial estimate and it serves to refine this initial pose estimate. Since the initial rotation angle is very close to its correct value, the rotation involved in the projection can be linearized around the initial estimate. More precisely, during the iterative operation of the EM algorithm, the estimate for the angle does not change extremely at each stage provided that its initial estimate is sufficiently good. Therefore, we can linearize the rotation operation around the most recent estimate of the angle at each stage. This leads to an affine projection model, which allows us to optimize the resulting objective function linearly.

Suppose we denote the most recent angle estimate, which is the linearization point at this stage, by θ_o and we use the following notation,

$$\cos\theta_o \triangleq \mu_o \quad , \quad \sin\theta_o \triangleq \upsilon_o \tag{4.2}$$

Within the assumption that $\theta \approx \theta_o$, $\cos\theta$ and $\sin\theta$ can be linearized around θ_o as

$$\begin{aligned} \cos\theta &= \cos\left((\theta-\theta_o)+\theta_o\right) \\ &= \cos(\theta-\theta_o)\cos\theta_o - \sin(\theta-\theta_o)\sin\theta_o \\ &\approx \mu_o - (\theta-\theta_o)\upsilon_o \\ &= K_o - \upsilon_o\theta \end{aligned} \tag{4.3}$$

where

$$K_o = \mu_o + v_o\theta_o \tag{4.4}$$

and similarly,

$$\begin{aligned} \sin\theta &= \sin\left((\theta - \theta_o) + \theta_o\right) \\ &= \sin(\theta - \theta_o)\cos\theta_o + \cos(\theta - \theta_o)\sin\theta_o \\ &\approx (\theta - \theta_o)\mu_o + v_o \\ &= K_1 + \mu_o\theta \end{aligned} \tag{4.5}$$

where

$$K_1 = v_o - \mu_o\theta_o \tag{4.6}$$

We now define β to be a 3-D vector,

$$\beta = \begin{bmatrix} \theta & t_x & t_y \end{bmatrix}^T \tag{4.7}$$

The preceding approximations lead to an affine projection model similar to that given in Eq. 3.13.

$$\eta_i = P(M_j,\beta) = M_j\beta + K = \begin{bmatrix} (-v_o x_j + \mu_o y_j) & 1 & 0 \\ (-\mu_o x_j - v_o y_j) & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ t_x \\ t_y \end{bmatrix} + \begin{bmatrix} (K_o x_j + K_1 y_j) \\ (-K_1 x_j + K_o y_j) \end{bmatrix} \tag{4.8}$$

where η_i represents the projection of the model feature into the image coordinate system, M_j denotes the model feature, specifically defined in matrix form to be able to define a

projection model linear in β, and β represents the pose of the object.

Since the linearity property is preserved in the new projection model, the EM iteration steps used for optimization needs only minor changes. Eqs. 3.40 and 3.41 used to solve the pose estimate and the weights in the maximization and expectation steps now become

$$\psi_\beta^{-1}(\widehat{\beta} - \beta_o) + \sum_i \sum_j w_{ij} M_j^T \psi_{ij}^{-1} \left(Y_i - (M_j \widehat{\beta} + K) \right) = 0 \tag{4.9}$$

and

$$w_{ij} = \frac{N((M_j \widehat{\beta} + K), \psi_{ij})}{\frac{B}{1-B} \frac{m}{W_1 W_2} + \sum_k N((M_k \widehat{\beta} + K), \psi_{ik})} \text{ for all } i, j \tag{4.10}$$

respectively, with β and M_j defined as in Eqs. 4.7 and 4.8.

Chapter 5

Processing Raw Data

The algorithmic components of the developed object recognition system can be decomposed into preprocessing, segmentation, feature extraction and classification. Conceptually, the act of classification, in the narrow context, consists of determining which of the models from the 3-D model library best matches the information extracted from the laser radar imagery.

This chapter discusses in detail the constituent subsystems of the model-based recognition system used to process the input raw data from the sensor to extract compact information that will be used in the matching procedure. In particular, the preprocessing, segmentation and the feature extraction steps will be presented respectively. The approaches used in each subsystem will be described together with their implementation details. Each module will be used to process the range data shown in Fig. 5-2, generated from the range truth, given in Fig. 5-1, in accordance with the procedure explained in Section 4.2. In this chapter, the sample image to be processed contains an army tank as the target. The corresponding video image is shown in Fig. 5-3. The results of processing this input image by each successive step of the recognition algorithm will be illustrated pictorially at the end of each section.

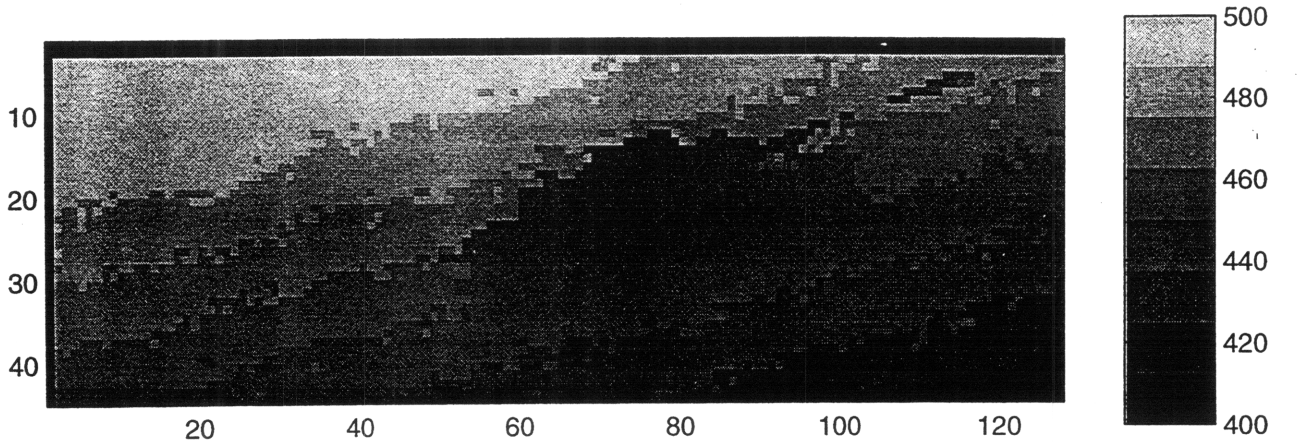

Figure 5-1: Range image of an M60 tank.

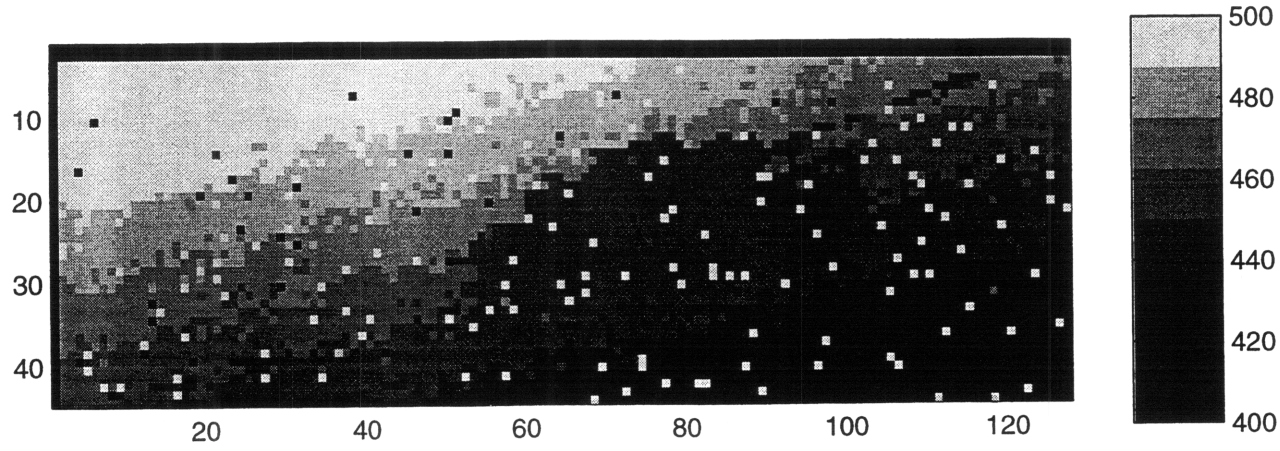

Figure 5-2: Range data of an M60 tank, artificially created from the range truth by addition of statistically independent, zero mean Gaussian noise to each pixel and random creation of anomalies.

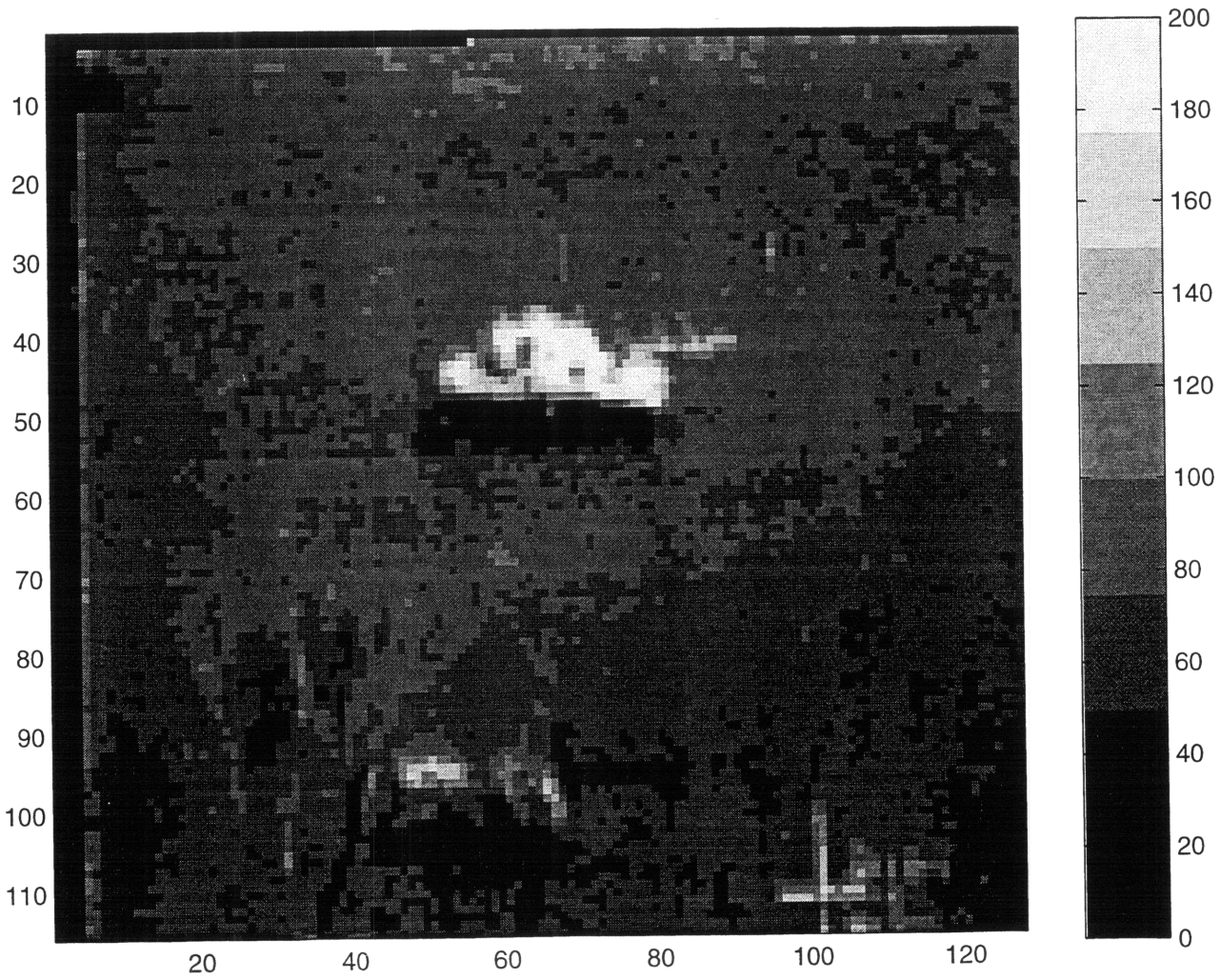

Figure 5-3: Video image of an M60 tank.

5.1 Preprocessing Step

The real laser radar range images, as discussed in Section 4.2, are characterized by coarse range precision, added noise and an appreciable amount of anomalous pixels. These images first need to be preprocessed to reduce these sensor-dependent effects. The purpose of the preprocessing operation is to improve the image quality so that the effectiveness of the subsequent processing steps is enhanced.

A wide variety of approaches can be used in the preprocessing step to accomplish this purpose. However, in this research, ad hoc image enhancement schemes such as median filtering have been avoided since they do not rely on the appropriate statistical model for the coherent laser radar range images. Ad hoc methods do not provide the quantitative performance criteria by which the impact of sensor capabilities on recognition performance can be assessed.

We have chosen to employ the multiresolution fast EM/ML algorithm as our preprocessor. The theoretical framework behind this algorithm was presented in Chapter 2. This approach is intended to overcome the degradation processes encountered in laser radar imaging process. It is distinguished from all other straightforward anomaly suppressing image enhancement techniques by the virtue of building on the sensor physics and thus providing the required, quantified, near-optimal performance characteristics.

The fast EM/ML algorithm has been developed from the conventional EM/ML algorithm using the special structure of the Haar-wavelet basis. This leads to a very computationally efficient and numerically robust method to find the ML estimates of real large imagery at various resolution levels subject to the anomaly suppression constraint. This approach, therefore, yields a compromise between resolution and anomaly suppression. The weights associated with the EM iterations can be used to determine the proper resolution at which adequate anomaly suppression is achieved, while preserving the fine-scale range truth details. This is called the 'stopping rule' and it is basically determining the resolution at which the fraction of the low weight pixels behaves according to the statistics of number of anomalous pixels in the image, which is a binomial random

variable whose mean is simply related to the easily-estimated CNR value [3].

The ML fits for the sample range image using 2×2-pixel and 2×4-pixel blocks can be seen in Figs. 5-4 and 5-5. As needed, the anomalous pixels have been successfully suppressed in all cases, with a very small number of exceptions in 2×2 (finest scale) case. The tank's body can be discerned at all resolutions. As the resolution increases, finer details, such as the barrel of the tank which is suppressed at low resolutions, become more evident. It has been shown in previous work that at high resolution, estimation performance approaches the ultimate limit set by the complete-data (CD) bound. Hence, although there may be a few unsuppressed anomalous pixels in the ML estimate, we have chosen to use the finest scale fit as the input to subsequent steps of the overall system.

5.2 Segmentation Step

A typical field-of-view of the laser radar might contain one or more objects of interest (targets) and one or more other objects (clutter), all of which are embedded in a background. Clutter refers to objects that are imaged, like buildings and trees, that are not the targets of primary interest. In the general context, the segmentation step is employed to distinguish all possible object regions, both target and clutter, from each other and from the background.

Clutter may dominate the imagery when the targets are sparse compared to the environment. However, for the case of framing mode laser radar range images, only a small area containing the target is imaged. Hence, the images used in this thesis do not involve clutter. Segmentation only serves the purpose of isolating the target from the background. However, this step is essential in the system since the target is partially embedded on the sloping ground on which it stands in the input image, as seen in Fig. 5-6, which represents the output of the preprocessing step. The subsequent feature extraction step determines the features by tracing the edge contours of the image. The algorithm which is used to locate the edge discontinuity curves in the image is incapable

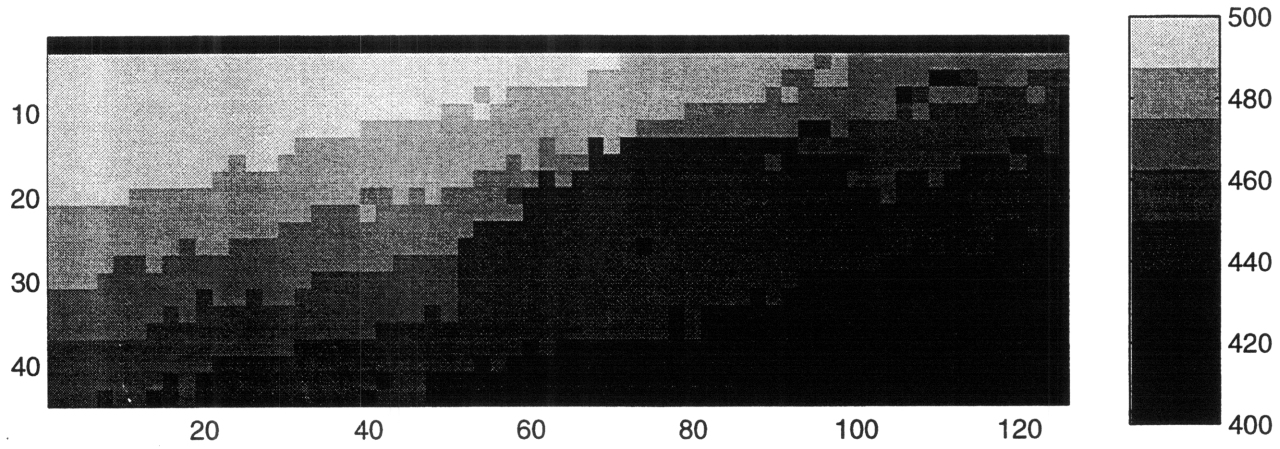

Figure 5-4: Multiresolution Haar wavelet EM/ML 2×2 fit to range data.

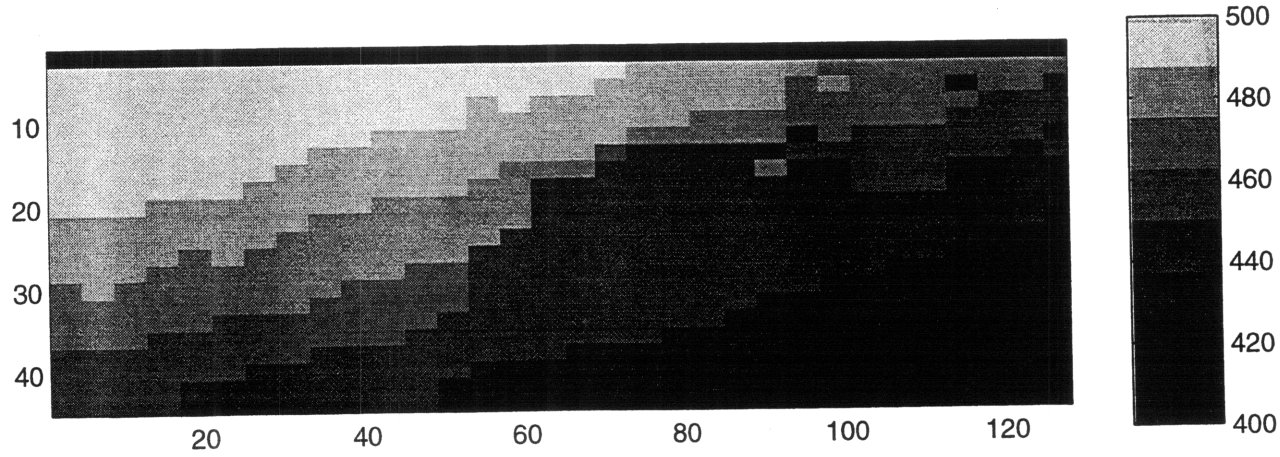

Figure 5-5: Multiresolution Haar wavelet EM/ML 2×4 fit to range data.

of determining the ground attachment of the target object in the range image. This results in great loss of information in terms of accurately modeling the object by its edge-based features. The features extracted in this way represent only the upper half of the object in the image. The edge curves determined by using the unsegmented image directly are illustrated in Fig. 5-7.

There are a number of techniques that can be used to solve this segmentation problem. Due to the nature of the input imagery, the simplest approach is to try to estimate the background, which is reasonably planar, in these images and to determine the pixels that lie above this plane which correspond to the target. For this purpose, planar range profiling is used to find the maximum-likelihood (ML) estimate of the background.

ML planar range profiling has been presented in Chapter 2. This method is employed to fit a planar surface to the input image neglecting the presence of the target and assuming that the true range values of the pixels comprise a plane. The algorithm treats the target object as a mass of anomalous pixels placed on a planar background profile. These pixels can be located using the final weights of the pixels, which are provided by the expectation step of the iterative EM algorithm. The weight of each pixel represents the conditional probability that associated pixel is not anomalous. Hence, the pixels corresponding to the target are the low weighted pixels and can be determined by comparing the weights against a certain threshold. The pixels whose weights are less than this threshold value are selected to determine the target region in the image. The threshold is selected to be 0.5. This is a reasonable approximation to determine the low weighted pixels since the pixel weights are clumped around 0 and 1. Figs. 5-8 and 5-9 illustrate the fitted planar background and the resulting segmented image. Note that all segmented images in this section are displayed on light backgrounds corresponding to pixels having much higher range values.

However, since thresholding is used to determine the pixels corresponding to the object, the algorithm skips an appreciable amount of pixels at the bottom of the object, whose range values are very close to those of the background. The coarse range resolution

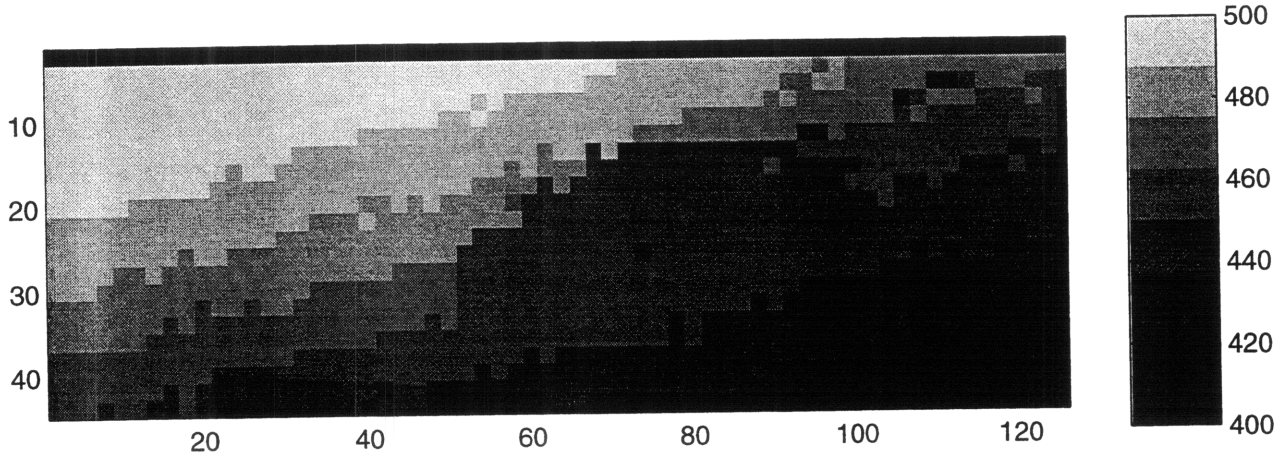

Figure 5-6: Input image to segmentation step.

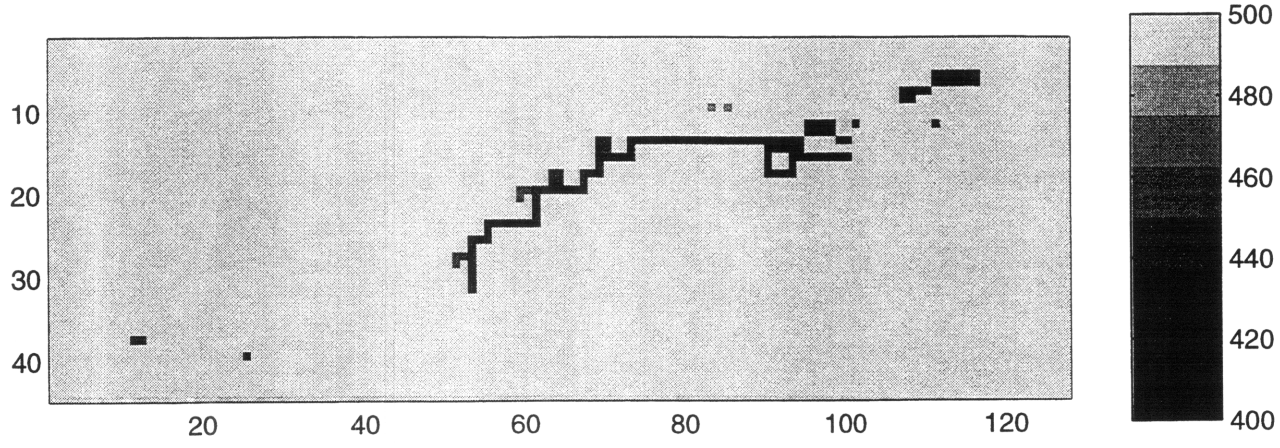

Figure 5-7: Edge discontinuity curves corresponding to the unsegmented image.

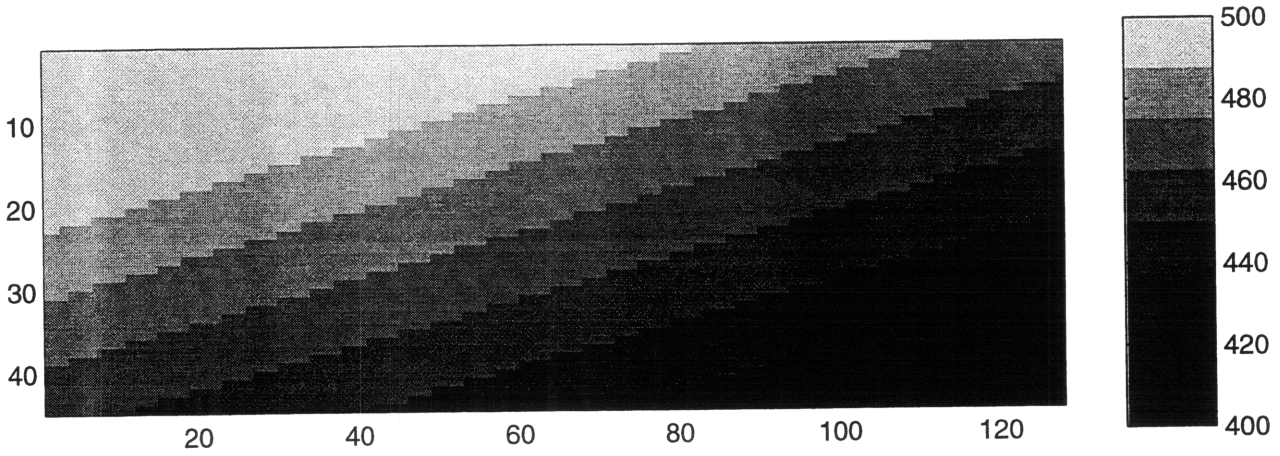

Figure 5-8: Planar range profile fitted to the range image.

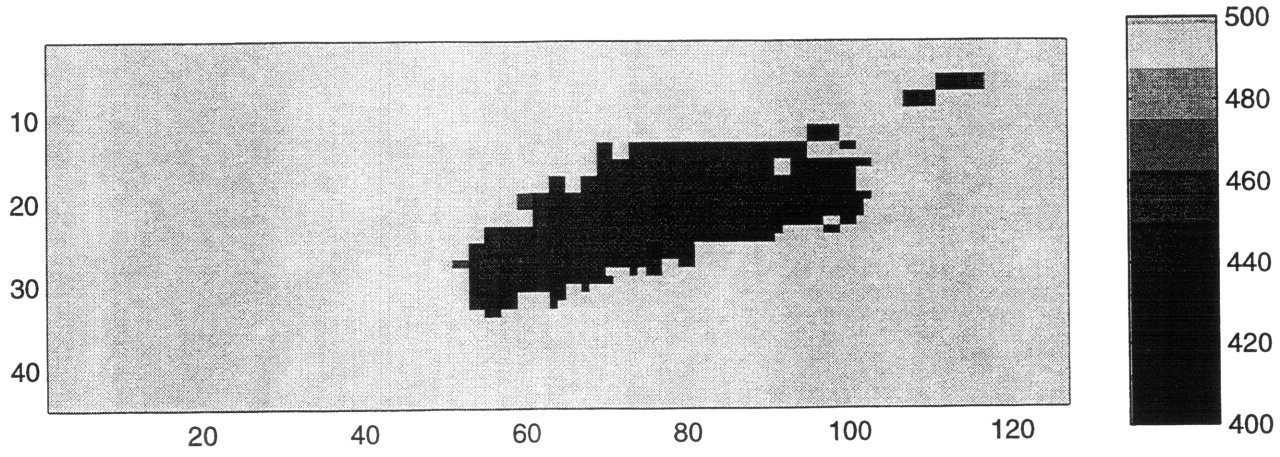

Figure 5-9: Pixels corresponding to the target determined by locating the low-weighted pixels in fitting a planar surface to the input image.

of the input imagery is another factor that makes it difficult to determine the ground attachment of the target.

The preceding segmentation procedure can be improved by using planar range profiling separately for the target and the background. The range values corresponding to the target and the background are assumed to constitute two different planes. The idea is to estimate the planar surface associated with the target and to compare this plane with the original image to determine the pixels consistent with it.

This target/background planar profiling can be achieved in a systematic manner using planar range profiling sequentially. The first step is, as described before, to fit a planar surface to the original input image and extract the anomalous pixels which correspond to the upper part of the target object by thresholding. This image is referred to as the first segmentation. These pixels contain adequate information to determine the plane associated with the target, which we will refer to as the target plane. Planar range profiling is used once more to estimate the target plane, using these pixels only. Fig 5-10 shows the resulting plane.

The next step is to compare the estimated target plane with the original image to extract the pixels whose range values lie on this plane. To make this comparison, a weight-based procedure is used. The weight of each pixel of the original image, which represents the conditional probability that the associated pixel is not anomalous, is evaluated conditioned on the assumption that the true range plane for this image is the target plane. The resulting high weighted pixels are determined by selecting those pixels that have weights greater than 0.9 corresponding to nonanomalous pixels which are on the target plane. A threshold of 0.9 is selected to be able to extract the truly nonanomalous pixels on the target plane. The pixels selected as a result of this procedure are displayed on a light background, corresponding to pixels with very high range values in Fig. 5-11. This image is referred to as the second segmentation.

The application of two stage segmentation is hindered by the very low resolution nature of the input image. The bottom of the tank in the sample image is embedded in

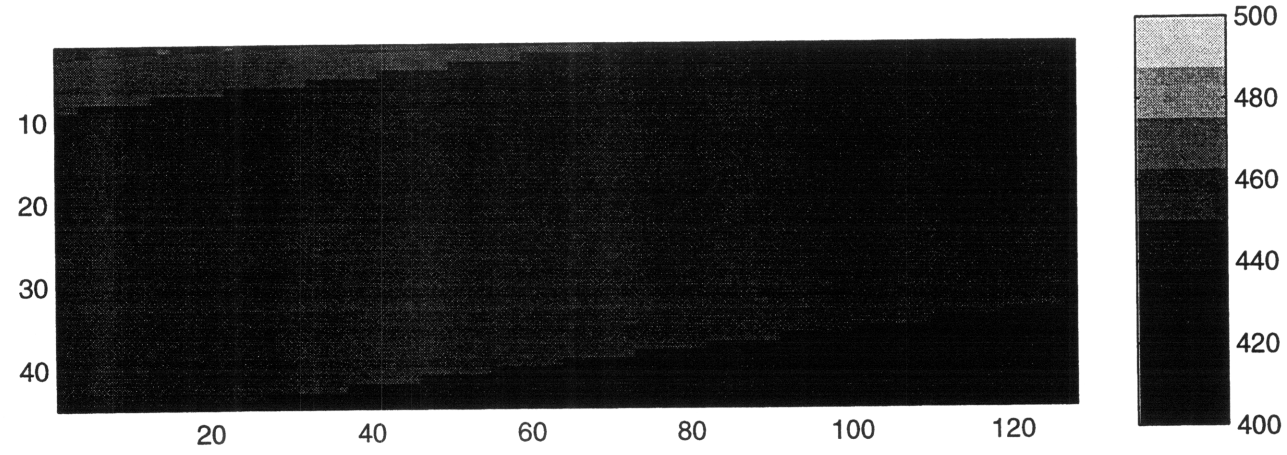

Figure 5-10: Estimated target plane.

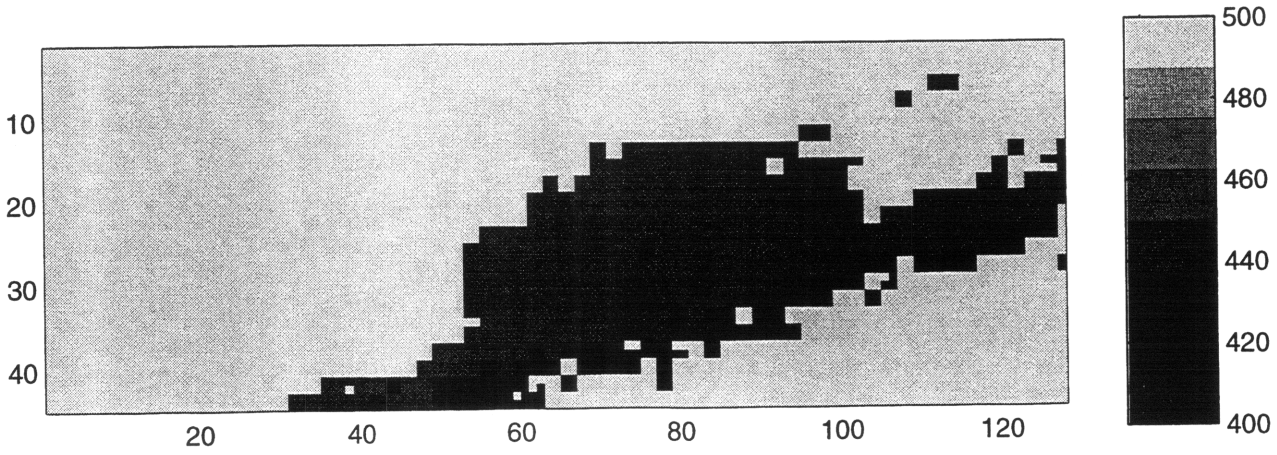

Figure 5-11: Pixels that lie on the estimated target plane.

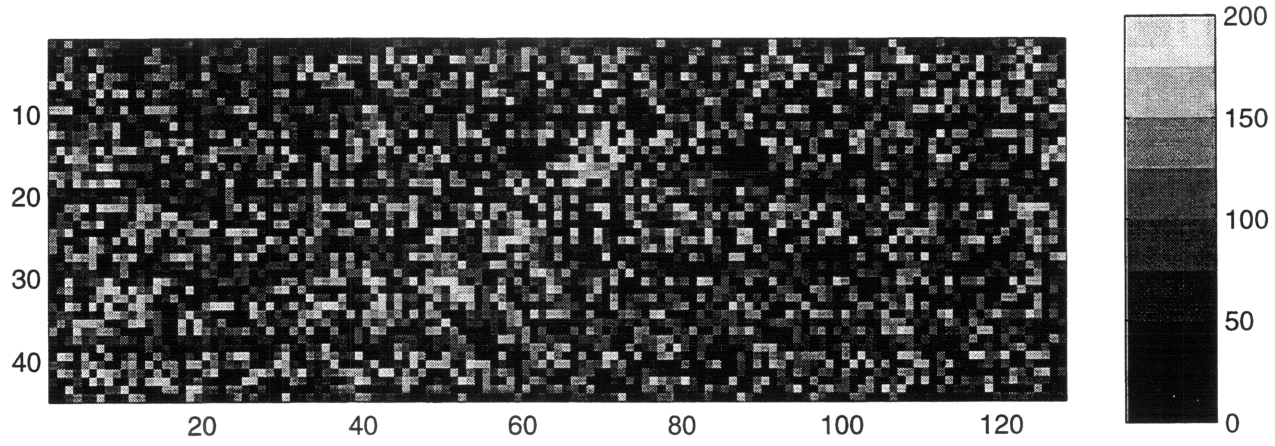

Figure 5-12: Intensity image.

the ground. The range values of the pixels at the bottom of the tank are very close to range values of the pixels on a tiny strip along the sloping ground, that the tank stands on. Therefore, the procedure described above extracts that tiny strip as being part of the tank since the range values of the pixels of that strip also lie on the estimated target plane. This phenomenon is illustrated in Fig. 5-11.

Fig. 5-11 shows us that the range images do not contain sufficient information to locate the ground attachment of the target precisely. This is an important issue, since accurately discerning the shapes of the targets in the raw data is essential to matching the object model to an object in the image. Use of the associated intensity image of the range image is hopeless since the reflectivity contrast of the target and the background is not considerable and the enormous intensity fluctuations due to speckle behavior even conceals the presence of a target in the image. The corresponding intensity image for the range image is seen in Fig. 5-12. The colorbar on the right indicates that the intensity values lie between 0 to 200 bins.

Under these circumstances, it is reasonable assume that the intercept of the line parallel to the strip, which defines the bottom edge of the tank, is a random variable uniformly distributed between the intercepts of the parallel lines defining the boundaries of the strip. The least squares estimate of a random variable, which minimizes the expected mean square error, is its mean value. For this case, this corresponds to placing the ground attachment on the line passing through the center of the strip. To recover the overall shape of the target, the front and the backward boundaries of the object are located in the first segmentation and extrapolated until they intersect the ground attachment line. To facilitate the implementation of this procedure, both the first segmentation and the second segmentation images are rotated to align the bottom of the tank with the horizontal pixel grid. This is achieved, for the image in Fig. 5-11 by a 15 degree clockwise rotation. The ground attachment line is determined as the line passing through the center of the strip in the second segmentation. The front and the backward boundaries of the object are determined using the first segmentation and are extended straight down until they hit the ground attachment line. The target region determined in the first segmentation is augmented by the pixels bounded by these lines in the second segmentation to constitute the overall target region in the image. The final segmented image is obtained by back rotating this image by the same amount. This process is demonstrated in Fig. 5.13. Fig. 5-14 displays the final segmented image on a light background corresponding to pixels having much higher range values. This image is used as the input to the subsequent feature extraction module.

It is shown in the recognition experiments in the next chapter that this approximation works fairly well for the targets of interest in the alignment process. However, difficulties may be encountered for curved objects since our algorithm estimates the bottom part of the object by sharp corners and straight lines.

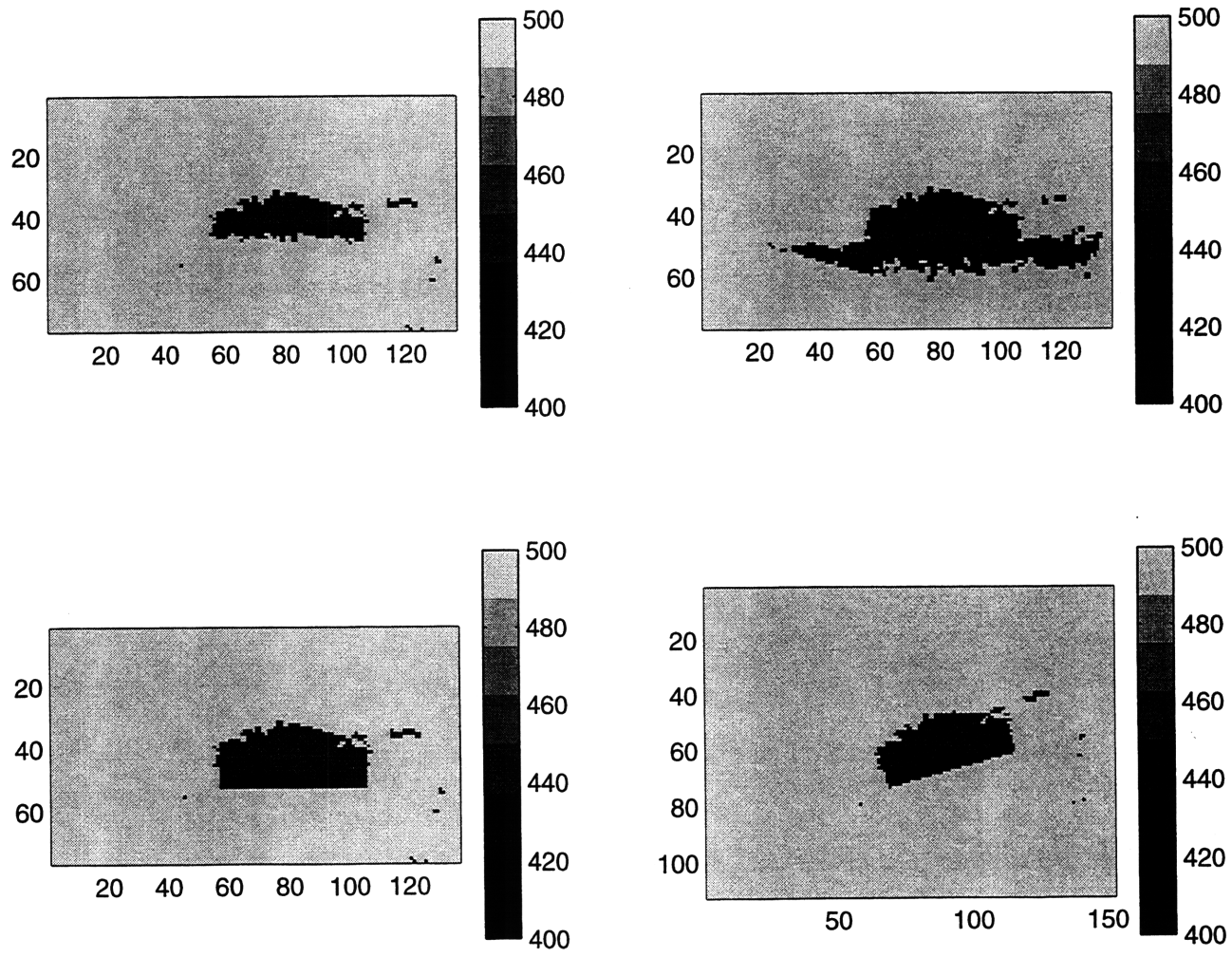

Figure 5-13: Final segmentation process: top figures illustrate the first and second segmentations rotated to aling the bottom of the tank with the horizontal pixel grid; the bottom figure on the left is the first segmentation augmented with the estimated additional target region; the bottom figure on the right, the final segmented image, is obtained by back rotating this image.

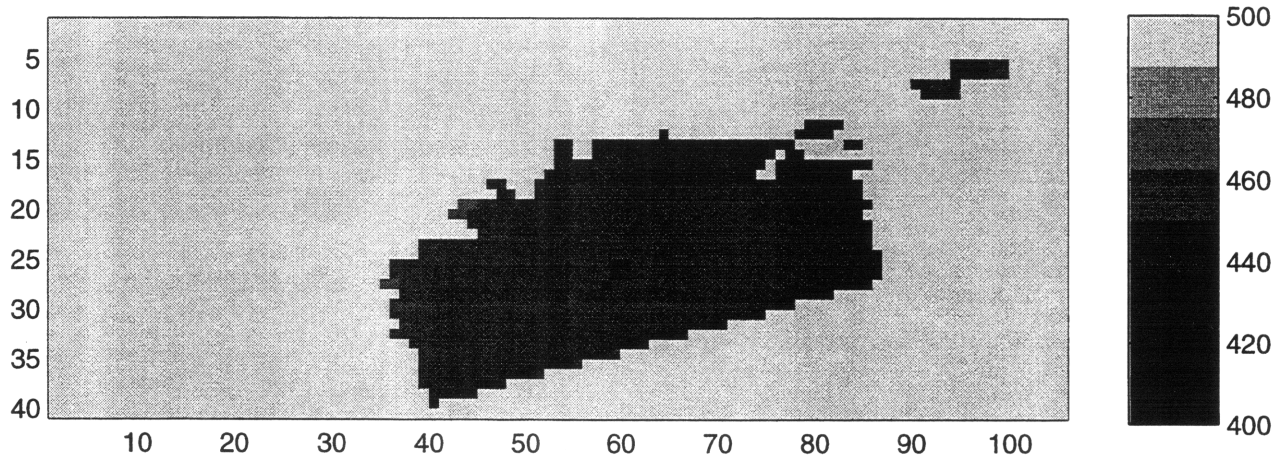

Figure 5-14: Final segmented image.

5.3 Feature Extraction

The object recognition process is crucially based on the matching of the image and the model data. Both the image and the object model are represented by a set of edge-based features and matching is performed in this new domain. Hence, the feature extraction step directly affects the classification performance of the matching algorithm.

The primary goal of the feature selection is to obtain compact information about the image and the object to be used in matching, which results in computational efficiency in the search process. This module is intended to extract the necessary information from the inputs of the recognition system to identify the target in the image. The quality of these features is essential for target classification since classification depends strongly on whether these adequately represent the target to distinguish different objects.

Edges contain a great deal of information about arbitrarily shaped objects. Use of edge-based features in object recognition has been proven to be effective in the past. Basically, the feature extraction module consists of two separate steps. The edge extractor

attempts to extract the range discontinuities that correspond to object boundaries in the input image. The following feature extractor decomposes the extracted edge contours into subcontours, on which the feature points are located. In the following sections, these subsystems will be discussed in more detail.

It is important to emphasize at this point that the same feature extraction process is applied to both the preprocessed, segmented range image and the rendered image generated from the object model. Using the same processing enables us to select the relevant features in a similar fashion for the target in both images. This aids the recognition processor, which attempts to locate model features in the image. However, since the two images correspond to target representations in different domains and thus are different in nature, determining the relevant features from the object model that are likely to be detected in the image is not always possible. It will be shown in the next chapter that the features extracted in this manner give good matching performance in recognition experiments.

5.3.1 Edge Extractor

The edge extractor tries to determine the boundaries of the objects in the image. These edges are accurate and compact representations of the overall shape of the objects.

Many different techniques can be used to locate the edge contours in an image. For the case of range data, edges may simply be regarded as borders separating the areas of different range values. This is also valid for the binary image corresponding to the rendered object model. Therefore to locate the edge contours, the simplest approach is to use thresholding between neighboring pixels.

The segmentation step extracts the target region and places the target object on a background having zero range values. Therefore the target is characterized by having high range values in the segmented image. The extractor finds the edge discontinuity curves by searching the image and thresholding the neighboring pixels. This is done by comparing each pixel to its 8 nearest-neighbor pixels. If there is a nearest-neighbor of

the center pixel, whose range value is less than that of the center pixel by more than a pre-chosen threshold, then the center pixel is selected to represent a range discontinuity, which constitute the edge contours at the object boundaries. Note that in thresholding process, the pixel is selected only if its range value is greater than that of its neighbor by an amount greater than the threshold. In this way, we only pick the pixels on the target boundary as feature points, not allowing the representation of the same edge behavior by two distinct pixels. The edge discontinuity curves corresponding to the sample image and the object model are shown in Figs. 5-15 and 5-16 respectively.

5.3.2 Feature Extractor

The next step in the feature extraction process is to decompose the determined edge curves into subcontours, which correspond to the extracted feature points and convey relevant information to construct the associated feature vector. This decomposition is done by segmenting the edge curves into smaller fragments of predetermined size.

This segmentation is achieved via a "march-down-the-curve" process. The search in the image is started: Whenever a pixel is found to be an edge discontinuity point, the curve through that particular point is traced upwards and downwards respectively and segmentation into groups of a certain number of pixels is done. This process goes on until all the curves are traced and segmented. Each of the resulting fragments is used to construct features consistent with the 2-D Point Feature model used. According to this model, the features contain information about the x and y coordinates of the extracted feature point. This information is determined using the curve fragments. Each curve fragment is assumed to represent a feature point whose coordinates are obtained as the mean value of the pixel coordinates of the segment. For image features, each feature point also preserves the range information corresponding to the average of the range values of pixels forming the fragment since this information is needed in mapping these features into the new coordinate system defined in Section 4.4.1.

The fragment length used in segmentation is arbitrary as long as it is small enough

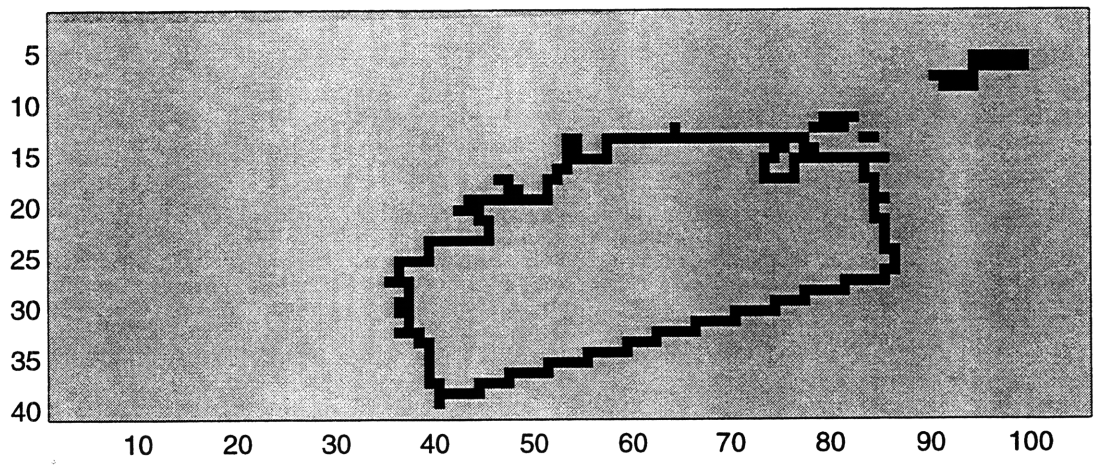

Figure 5-15: Edge discontinuity curves corresponding to the segmented range image.

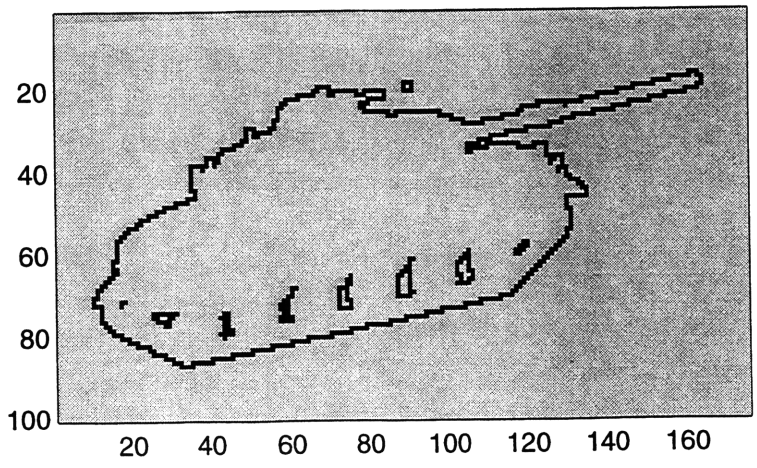

Figure 5-16: Edge discontinuity curves corresponding to the rendered image of the object model.

so that the set of the features constitute a good representation of the target. A different value of segment length has been used for the range image and the rendered image, since the two images have different sizes and the target region in these images contain different number of pixels. The segment length used in each image is determined by the constraints imposed by the covariance structure used in the objective function of the recognition module, which will be discussed in Section 6.1. The extracted feature points from the range image and the object model are illustrated as points on the located edge curves in Figs. 5-17 and 5-18.

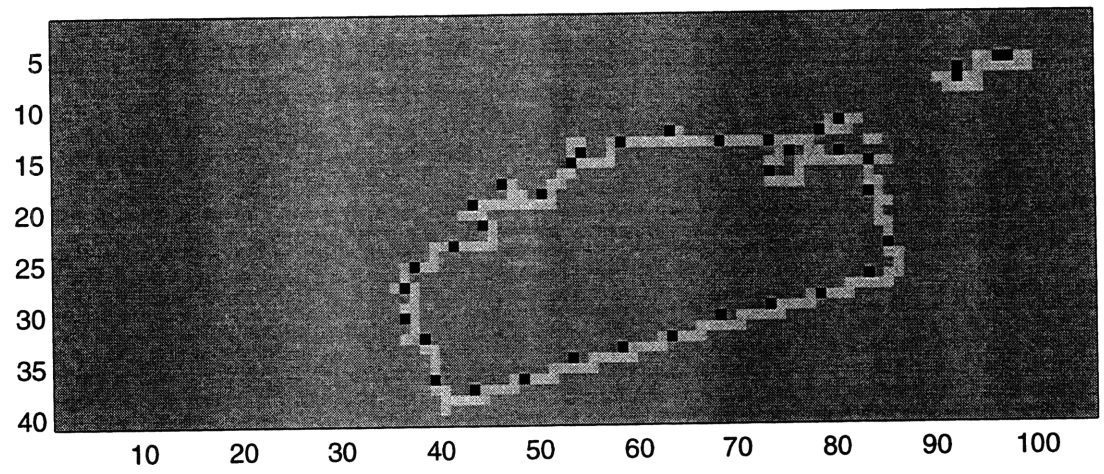

Figure 5-17: Extracted feature points from the range image located on the edge curves.

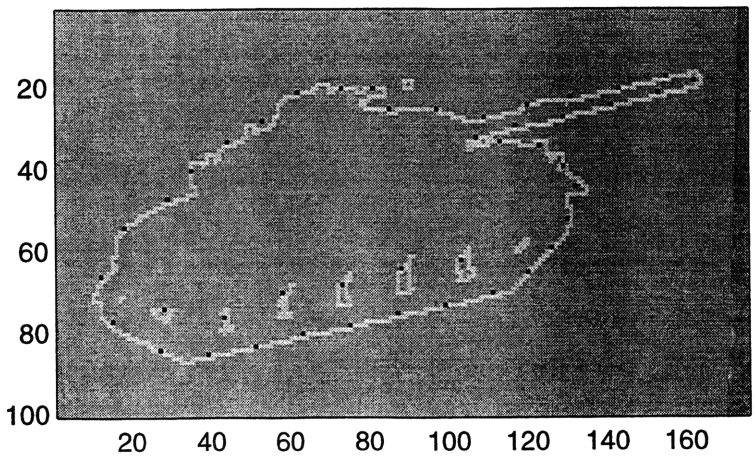

Figure 5-18: Extracted feature points from the rendered image located on the edge curves.

Chapter 6

Pose Estimation and Classification

The matching step constitutes the most important part of our system. Essentially, our primary goal in the overall recognition system can be stated compactly as finding and evaluating the alignment of the image and the model data. The parameter to be estimated in finding the correct alignment is the pose of the object in the image. This step estimates the pose of the target in the image and performs matching between the image and the model features. The output is a score which gives an indication of the degree of alignment between the image and the particular object model. This procedure is applied to all models in the data library that account for the possible targets and the resulting scores are compared to detect which of the models correspond to the target in the image.

In this chapter, we present the alignment results between the sample image we have been processing so far and some representative models selected from the data library. We start by examining the objective function used in pose estimation and scoring. First, the approach used in determining the parameters involved is explained . The behavior of the objective function is investigated using probes, which are the samples of the objective function passing through the known location of the object in the pose space. The alignment results and the scores of alignments, which lead to classification of the target in the image, are then presented. The results are analyzed by looking at multiple trials of matching simulated range images randomly generated from the range truth with the

correct model and one of the other models and by examining the effect of resolution on the performance of the system. Finally, system performance is examined as a function of the sensor parameters.

6.1 Determination of the Required Parameters

The matching mechanism of the system employs the posterior marginal pose estimation (PMPE) method of Section 3.2.5, which is used in estimating the pose of the object in the image and acts as a scoring measure for determining the correct model. The PMPE objective function was given in Eq. 3.36. We assume that no pose prior is available and the first term can be left out. The final form of the objective function is therefore

$$L(\beta) = \sum_i \ln \left[1 + \sum_j \frac{W_1 W_2}{m} \frac{1-B}{B} N(Y_i; (M_j\beta), \psi_{ij}) \right] \tag{6.1}$$

In Eq. 6.1, Y_i and M_j represent the image and the model features respectively. From this expression it is clear that the objective function is a measure of degree of alignment between the image features and the projected model features.

Even after the choice of the statistical models and formulation, it is still necessary to supply the specific parameters for the model, namely the background probability, B, and the covariance matrices of the normal densities, ψ_{ij}, to be able to use this objective function in the required optimization.

The background probability may be estimated by taking simple statistics on images from the domain. For range imagery, the background features may come from other objects present in the image, from anomalies which could not be suppressed by the front end processor or from anomalous pixels that arise due to segmentation step. For convenience, the background probability is assumed to be constant for all image features. The proportion of the extracted background features is quite small in comparison with the total number of the features for the available data set. Therefore, the background probability is set to 0.1 throughout the recognition experiments.

The covariance matrix that appears in the statistical model of the matched image features is allowed to depend on both the image feature and the model feature involved in the correspondence. However, substantial simplification results by assuming that the covariance matrix is independent of the model features and that the feature statistics taken relative to the coordinate systems attached to each image feature are stationary in the image. This approach, known as "oriented stationary statistics" [18], permits estimation of the covariance matrix using observations on matches done on a sample image and the associated rendered object model. During this process, the pose of the object is kept the same in the two images. Correspondences are made between image features and model features. Suppose that the x and y coordinates of the observed image and the model features are given by the vectors Y_i and $\widehat{Y}_i$, respectively. The observed residuals,

$$\Delta_i = Y_i - \widehat{Y}_i \tag{6.2}$$

are transformed into coordinate systems specific to each image feature defined by the local edge at the feature point by means of the corresponding rotation matrices, R_i,

$$\Delta_i^{'} = R_i \Delta_i \tag{6.3}$$

Then, the stationary covariance matrix of the matched feature fluctuations observed in the feature coordinate systems, is estimated using the maximum-likelihood method as follows,

$$\widehat{\psi} = \frac{1}{n} \sum_i \Delta_i^{'} \Delta_i^{'T} \tag{6.4}$$

During the recognition process, the covariance structure representing the fluctuations of image feature coordinates is required. Therefore, the constant covariance matrix found in the image feature coordinate systems needs to be specialized to each image feature by transforming it back to the image system using the rotation matrices, R_i,

$$\psi_i = R_i^T \widehat{\psi} R_i \tag{6.5}$$

Previous studies on the structure of the covariance matrix reveals that the equiprobable contours associated with the covariance matrix should be elliptical. This is expected since the variance pertaining to feature deviations perpendicular to the edge contour is different from the variance related to deviations parallel to the edge contour. However, it is possible to achieve a circularly symmetric covariance structure by adjusting the parameters used in feature extraction process considering the mechanisms resulting in feature fluctuations. Using a circularly symmetric covariance structure removes the need for the rotation step, hence improves the computational efficiency.

Deviations perpendicular to the edge contour are due to sensor characteristics and the edge detection algorithm. The perpendicular variance, which is assumed to be stationary for the image features, can be estimated experimentally using the procedure explained above. On the other hand, deviations parallel to the edge contour are affected primarily by the feature spacing used along the contour. This assumption is reasonable since using 20-pixel spacing instead of 5-pixel spacing in extracting the features results in a much larger variance along the contour. Therefore, it should be possible to adjust feature spacing so that a circular covariance structure can be obtained for feature fluctuations.

Suppose that the features are extracted using d-pixel spacing along the contour. This suggests that the predicted location of each of the image features can be at most $\pm\frac{d}{2}$ pixels away from the true location. This is illustrated more in Fig. 6.1. If we assume that the starting points in feature extraction are equally likely, it is reasonable to model the parallel deviation of a feature to be a uniform random variable over the $\frac{d}{2}$-pixel neighborhood of the true feature location . Hence, the variance along the edge can be taken to be equal to the variance of this uniform random variable given by

$$\sigma_{\parallel}^2 = \frac{d^2}{12} \tag{6.6}$$

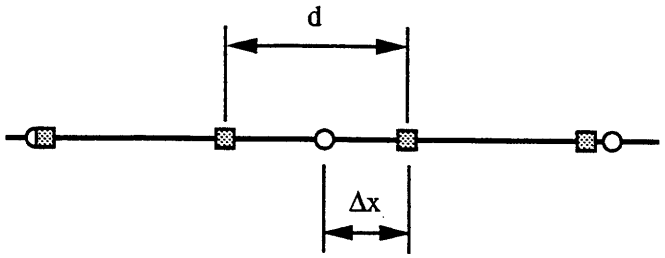

Figure 6-1: Effect of feature spacing on the variance of features along the contour; $|\Delta x| \leq \frac{d}{2}$, where Δx is the error between the projected model feature location and the actual feature location and d is the model feature spacing.

However, the starting point of feature extraction process is not equally likely and therefore, it is reasonable to model the parallel deviation of a feature to be a Gaussian random variable centered on the true feature location with the same variance as the uniform density, $\frac{d^2}{12}$.

This result suggests that the feature spacing can be selected properly to achieve a circularly symmetric structure. In particular, we choose the feature spacing as

$$d = \sqrt{12\sigma_{\perp}^2} \tag{6.7}$$

where $\sigma_{\perp}^2$ is the estimated variance for the perpendicular deviations. The resulting covariance matrix then becomes

$$\psi_i = \sigma_{\perp}^2 I \qquad \forall i \tag{6.8}$$

The approach explained above has been used to determine the covariance matrix used

in our recognition experiments. First, the variance for deviations perpendicular to edge contours has been estimated by the maximum-likelihood method using a sample image from the domain and a 2-D view of the object model rendered with the same pose. The value of $\sigma_\perp$ has been found to be 0.243 meters. This value has been used to find the feature spacing, d, to be equal to 0.8421 meters using Eq. 6.7. The equivalent value in terms of pixels is found for the 2-D views of the correct object model to be 12 pixels and used in feature extraction for all of the models. The pixel spacing value used for the image features is chosen so that the resulting features form an adequate representation of the target edge discontinuity curves.

6.2 Probes of the Objective Function

Pose estimation is the first step in the matching process between the image and the model data. The matching system estimates the pose of the target in the image using the PMPE method of Section 3.2.5. The statistical formulation used in this method leads to an objective function given in Eq. 6.1, which is repeated here for convenience,

$$L(\beta) = \sum_i \ln \left[1 + \sum_j \frac{W_1 W_2}{m} \frac{1-B}{B} N(Y_i; (M_j\beta), \psi_{ij}) \right] \tag{6.9}$$

The previous section presented our method of determining the required parameters involved in this expression. These predetermined values for these parameters will be used throughout our recognition experiments.

The pose estimation step requires that this objective function be optimized in the pose space. Before doing so it is worthwhile to investigate the behavior of the objective function by taking its samples at discrete locations in the pose space to find out if it has a peak close to the correct pose of the target in pose space. The resulting graphs are referred to as the 'probes' of the objective function.

Although rigid body motion is characterized by six degrees of freedom, we have reduced our search space to three dimensions. As explained in detail in Section 4.4, our

pose space is characterized by translation in x direction, denoted by t_x, translation in y direction, denoted by t_y, and in-plane rotation, denoted by θ. The correct pose of the target in the image with respect to the initial location of the object model is not known apriori. Therefore, the probes can only be used to illustrate the existence of peaks of the objective function along each component in pose space. The pose estimate found as a result of pose estimation experiments that will be explained in the next section is given below to demonstrate that the peaks occur very close to the estimated values.

$$\beta = \begin{bmatrix} \theta & t_x & t_x \end{bmatrix} = \begin{bmatrix} 0.0612 & 1.0631 & 3.5062 \end{bmatrix} \tag{6.10}$$

Samples taken along a line through the location of the true pose in pose space parallel to t_x axis are shown at the top in Fig. 6.2. Samples are taken at 0.5 meter intervals in [0,4] meter interval, which includes the correct value of the relevant pose parameter. The discrete sample points are joined with straight line segments for clarity. Similar probes have been taken parallel to t_y axis with a sampling interval of 0.5 meters in [0,4] meter interval, and θ axis, with a sampling interval of 0.01 radians in $[0, 0.1]$ radian interval, as illustrated in the same figure.

These probes demonstrate that the objective function has a prominent, sharp peak very close to the estimated location in the pose space. Note that in the general sense this method does not provide us with the optimum pose parameter values since a global searching was not performed. On the other hand, it does show that a reasonable pose estimate can be obtained via a search in the pose space. The next step is to perform this search, starting from a good initial value, by means of an appropriate numerical optimization algorithm.

6.3 Matching

The input to the matching subsystem consists of compact descriptions of the image and the model in the feature domain. With this data, the matching step deduces the identity

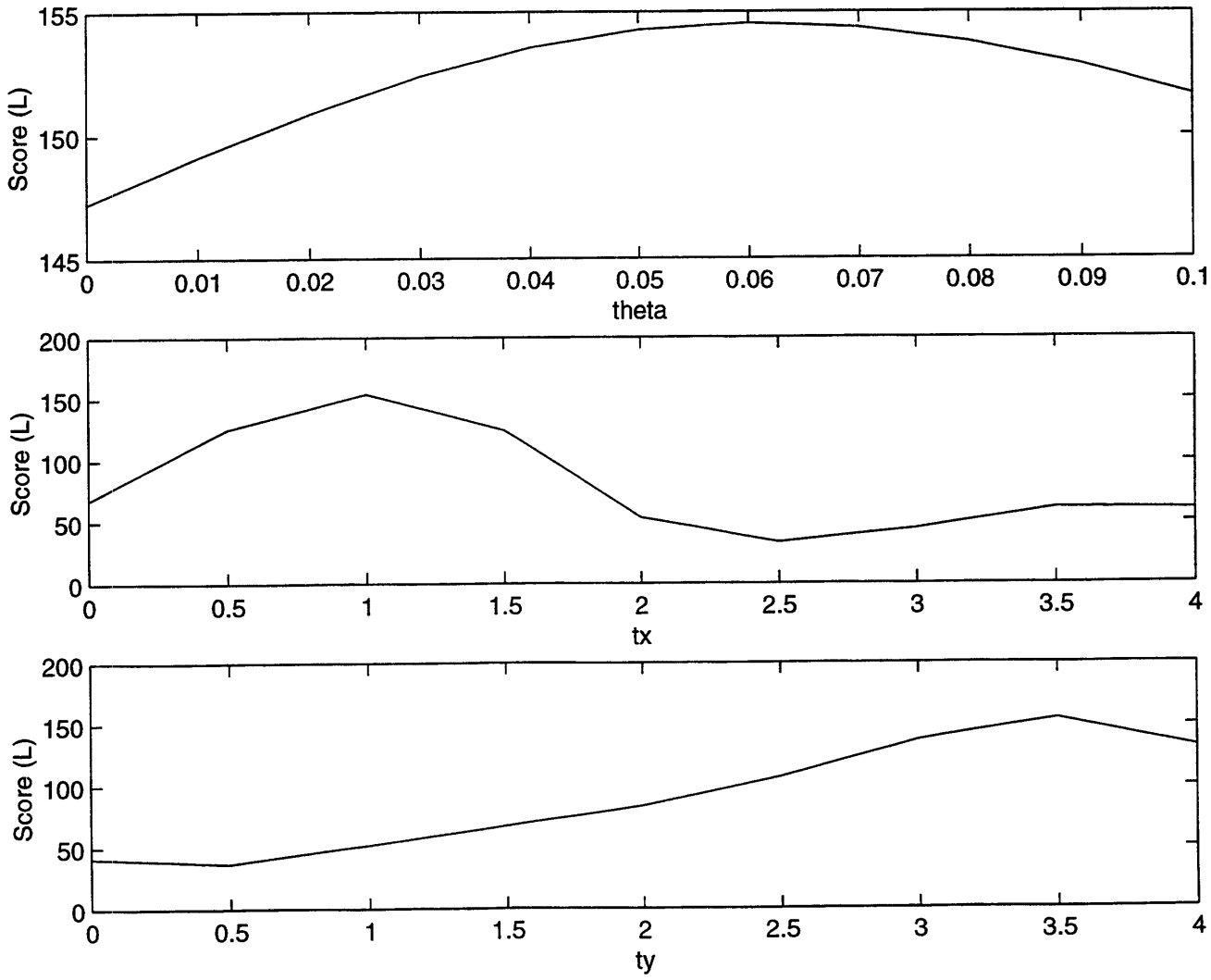

Figure 6-2: Probes along t_x , t_y and *theta* axes.

of the target in the data library that best matches the image of interest.

The matching module consists of two stages. In the first stage, the pose of the object in the image is estimated such that the maximum number of optimal pairings between the two sets of features and thus the best alignment between the two is achieved. In the second stage, the actual image features are compared with the model features projected onto the image plane with the pose estimated by the posterior marginal pose estimation (PMPE) algorithm as the measure for comparison. The alignment with each model is evaluated in this manner to classify the target in the image as one of the models in the data library.

6.3.1 Pose estimation

In this section, we present alignment results pertaining to features derived from the preprocessed, segmented sample image, which contains the M60 tank as the target. The resulting image features are sketched as black points in Fig. 6.3. Note that in this figure, each pixel has 0.1 meter width and height.

Four representative models have been selected from the data library to perform matching with the image of interest. Fig. 6.4 illustrate the rendered 2-D views of these models.

The top left rendered view corresponds to the actual target present in the image, namely an M60 tank. We have also included another tank model, the T80 tank model, to test the ability of our algorithm to distinguish different models of the same type of vehicle. Note that the tank models and the GMC CCKW truck model are almost the same size, i.e., 7-8 meters in length, whereas the Ford GPA jeep model is much smaller, i.e. 2.9 meters in length. This helps us emphasize one of the most important characteristics of our alignment procedure, namely performing matching using actual dimensions of the targets. These models are also transformed into the feature domain using the same feature extraction model. The corresponding features are plotted as dots on the edge curves in Fig 6.5. Table 6.1 provides data pertaining to the feature extraction process for the sample range image and the four test models used in the recognition experiments.

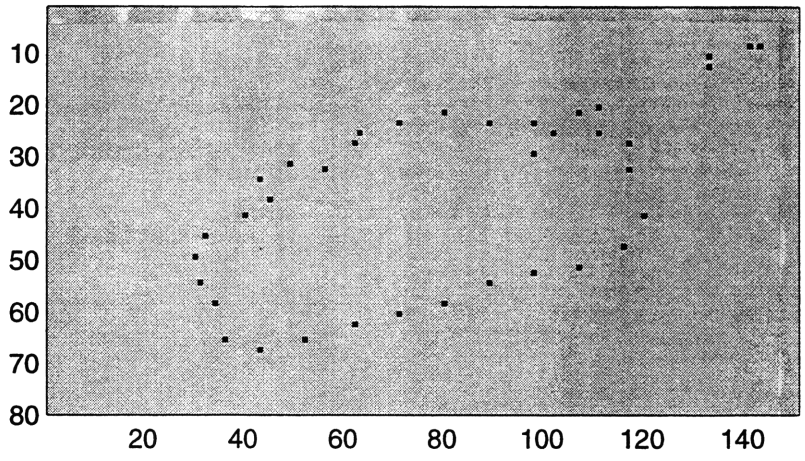

Figure 6-3: Image feature points.

	Feature Spacing	Number of Features
Range Image	5	37
M60 Tank Model	12	44
T80 Tank Model	12	38
GMC CCKW Truck Model	12	39
GPA Jeep Model	12	35

Table 6.1: Data associated with the feature extraction process.

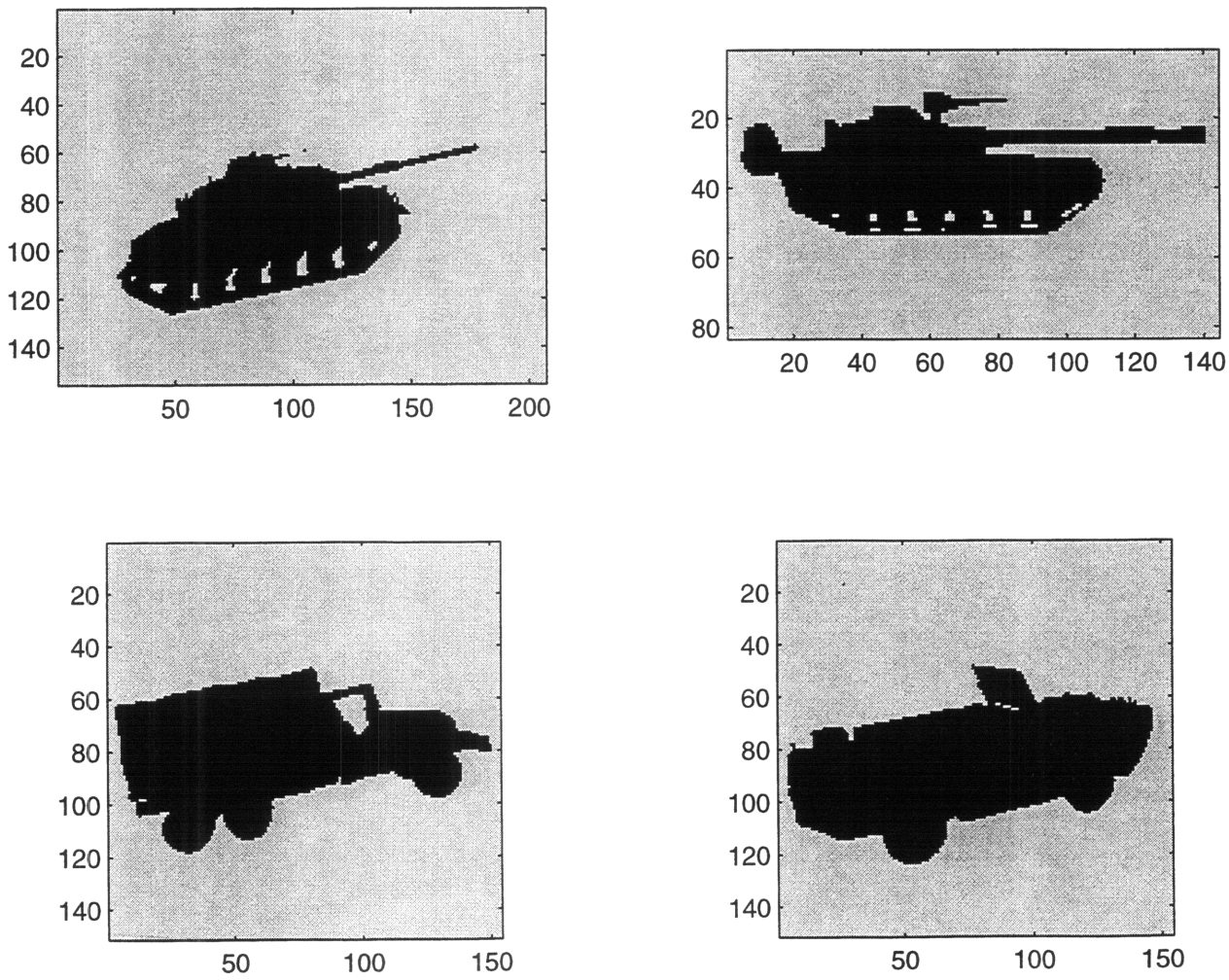

Figure 6-4: Rendered views of object models.

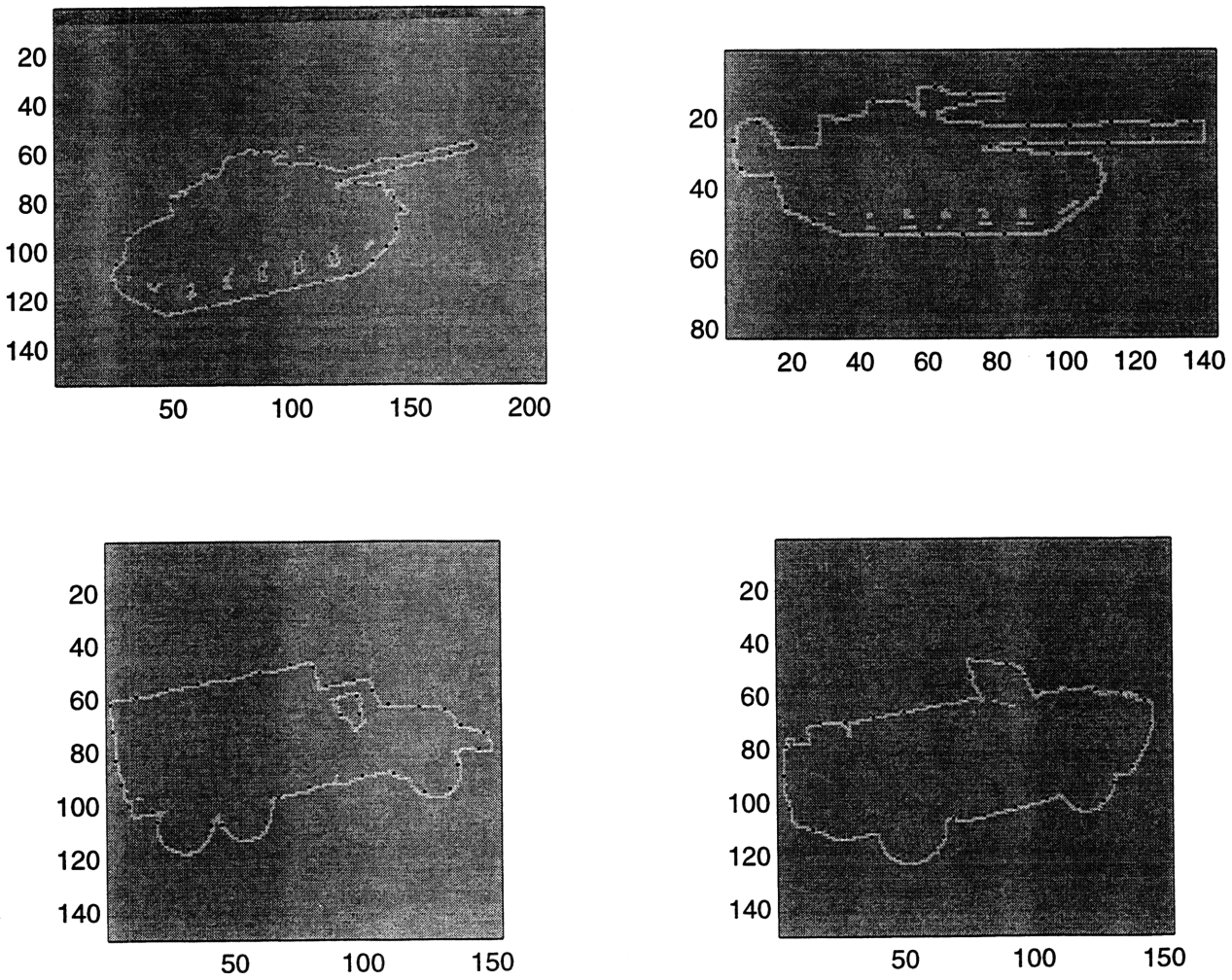

Figure 6-5: Model features.

Given the two sets of features and the necessary parameters, the PMPE method with the EM algorithm is employed to find the optimum alignment between the image and each of the models separately. As discussed in Section 2.2.1, the EM algorithm will converge to a local likelihood maximum. It requires proper initialization to reach the global likelihood maximum. Therefore in our research, we assumed that we are provided with a good initial pose estimate and our purpose is to refine the value of the pose vector by a local search in the pose space. In our experiments, the initial pose estimate is determined as a result of alignment by hand.

We start with the alignment of the image with the correct model, the M60 tank model. Note that in all figures in which alignment between two sets of features are displayed, the white dots represent the model features whereas the black dots represent the model features. Fig. 6.6 shows the initial alignment between two sets of features at the top, i.e., the alignment of the image with the rendered view. The pose of the object in the initial rendered views is referred to as the null pose, $\beta = [0, 0, 0]$, and is taken as the reference in evaluating the pose estimates. The initial pose is selected by an alignment done by hand. Fig. 6.7 illustrates the alignment of the image features with the model features projected onto the image with the pose estimated by the PMPE method using EM algorithm.

Since we do not know the actual pose of the target in the image with respect to the null pose, we cannot deduce how close the estimated pose is to the correct one. However, the resulting alignment is good enough in terms of aligning the boundaries of the object from which we can infer that we achieved a reasonably good pose estimate. The features do not perfectly match in this particular alignment. There are several reasons for this behavior. The most significant mismatch is observed at the corners of the object and on the barrel. The corner behavior arises due to characteristics of our processing mechanisms. The extracted image features tend to lie on sharp corners, at the top due to block structure of the preprocessor, the fast EM/ML algorithm, and at the bottom due to straight line extrapolation of the segmentor. The barrel mismatch is due to the fact that the elevation

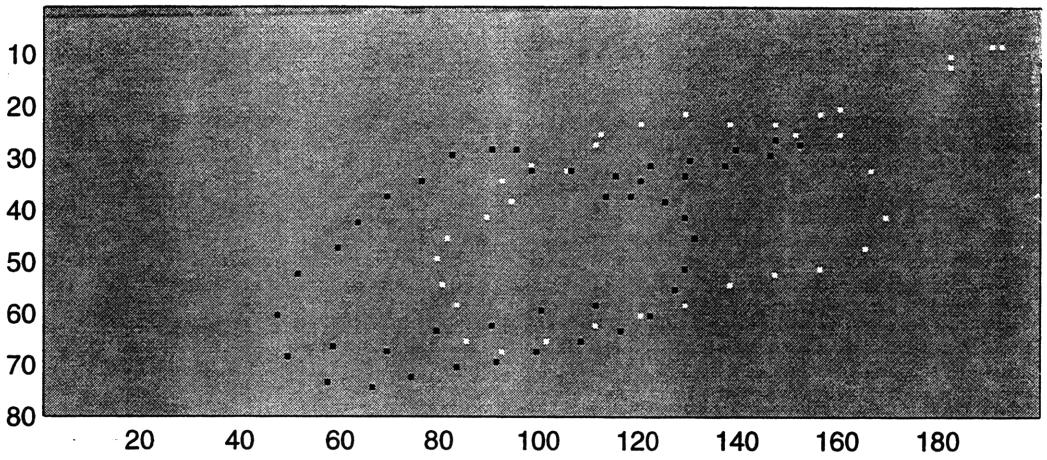

Figure 6-6: Initial alignment with M60 A3 Tank model.

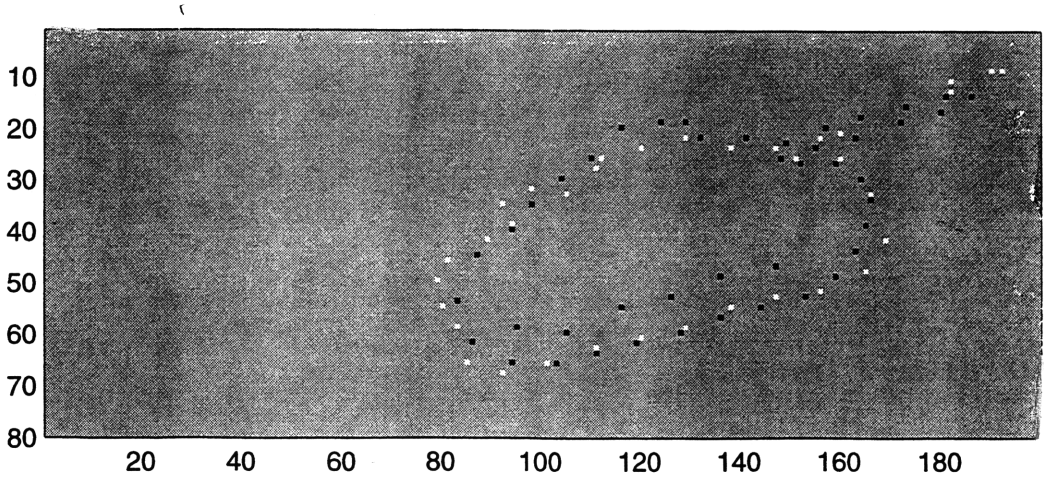

Figure 6-7: Final alignment with M60 A3 Tank model.

angle of the barrel may not be the same in the image and the model. This implies that it is difficult to deal with movable parts of the targets by our algorithm.

Nevertheless, the resulting alignment reveals that the overall shape of the target in the actual and predicted images match reasonably. This will be reflected in the value of the alignment score, which will be given in the next section in the context of classification of the target among possible candidates from the data library.

The initial and the EM alignments corresponding to T80 tank model are illustrated in Figs. 6.8 and 6.9 respectively. The final alignment results for the M60 tank model and the T80 tank model, illustrated in Figs. 6-7 and 6-9, show that there is a stronger match between the image data and the M60 tank model claiming that the target is more likely to be an M60 tank. This observation will also be reflected in the alignment scores. It is easy to understand the behavior of the pose estimate for the T80 tank model. The pose is estimated such that the maximum number of matches is achieved between two sets of features. The algorithm tries to match regions of the image and the model, where there is a cluster of features. This is revealed in the final EM alignment in Fig. 6.9, since the region in front of the tank close to the barrel, which isrich with features in both the image and the model, are aligned by the algorithm. The effect of using actual dimensions in the alignment is reflected in the same figure. Due to the size difference, the algorithm is incapable of matching all boundaries, instead it leans the object model to the front border of the tank in the image to get the maximum number of pairings between features.

Similar behavior can be observed in the alignment results of the truck model, illustrated by initial and EM alignments in Figs. 6.10 and 6.11. The truck model contains a similar region clumped with features in the front. The pose is estimated in such a fashion as to align this region with the cluster of image features. Also, the rotation angle is selected so that the boundaries of the truck are aligned with those of the tank in the image.

The last matching is performed with the jeep model. The initial and EM alignments

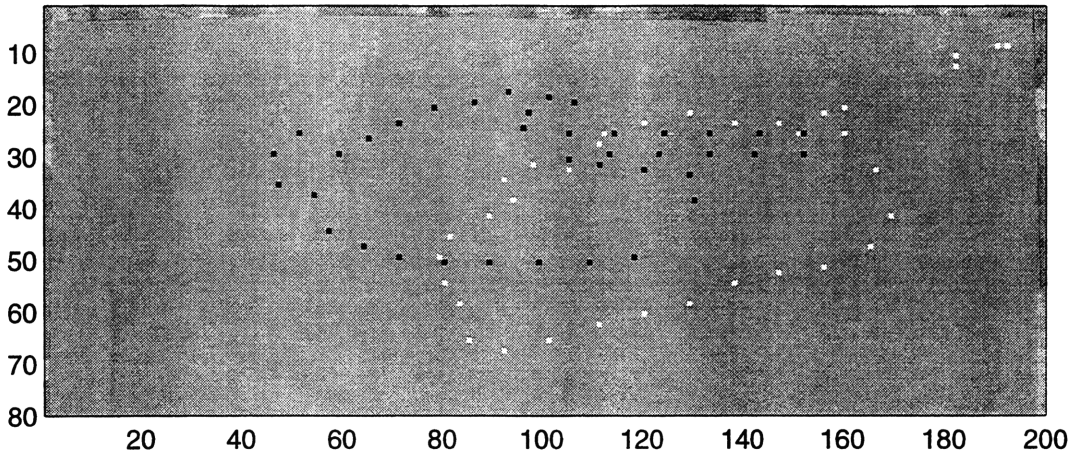

Figure 6-8: Initial alignment with T80 Tank model.

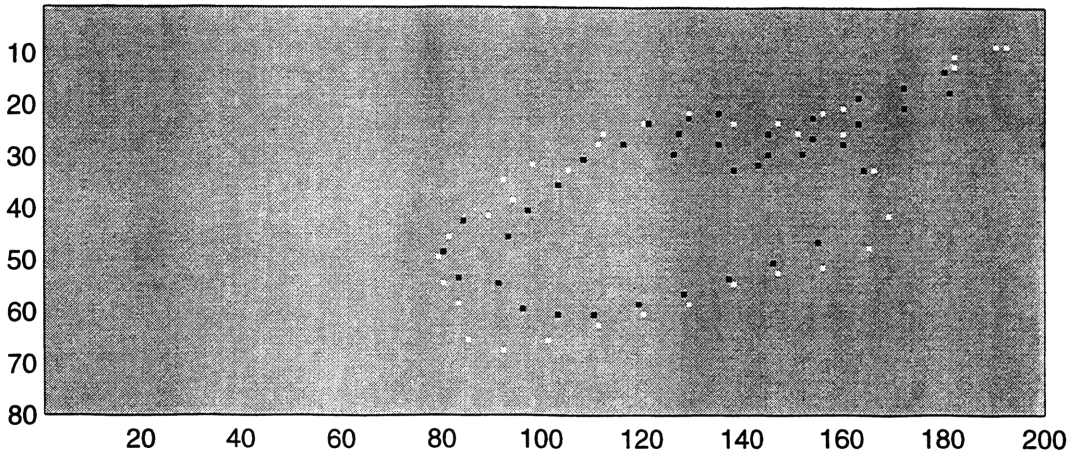

Figure 6-9: Final alignment with T80 Tank model.

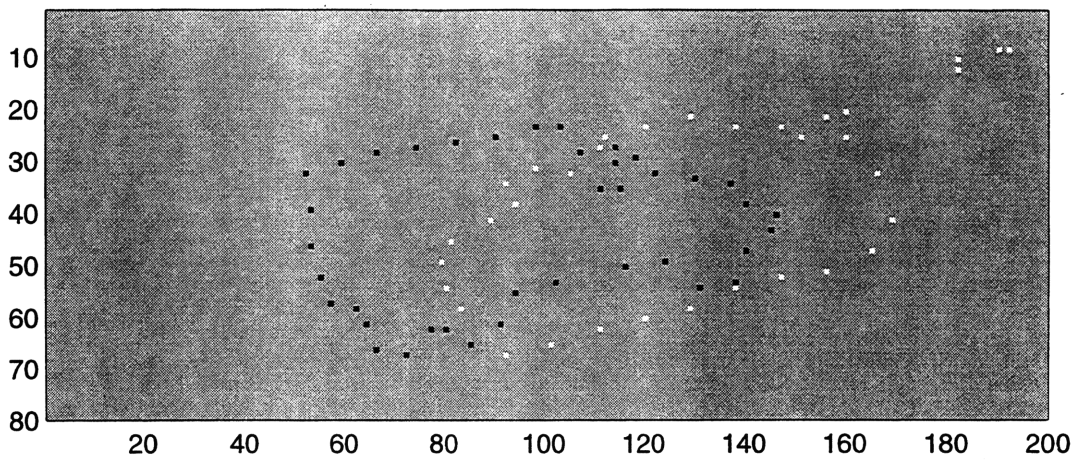

Figure 6-10: Initial alignment with GMC CCKW 353 Truck model.

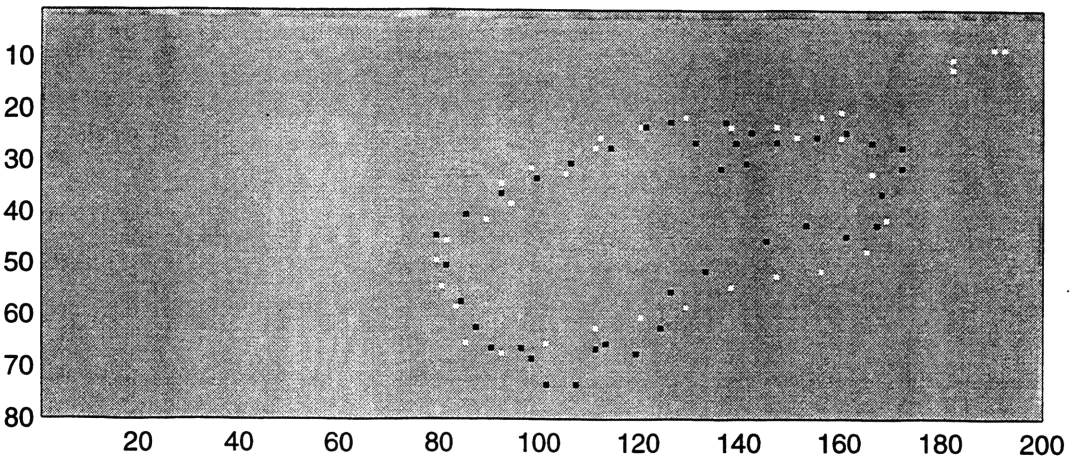

Figure 6-11: Final alignment with GMC CCKW 353 Truck model.

	Pose Estimates	Number of Iteration Steps
M60 Tank Model	0.0612 1.0631 3.5062	10
T80 Tank Model	0.2956 −0.2238 3.5444	6
GMC CCKW Truck Model	0.2551 0.1425 2.7671	25
GPA Jeep Model	0.5244 0.2038 −0.5451	12

Table 6.2: The EM pose estimates and the required number of iteration steps.

can be seen in Figs 6.12 and 6.13. This matching is a clear manifestation of the main characterization of our algorithm: a model whose size is very different from the true object is forced to have a low scoring. The best EM alignment depends extensively on the initial pose. Starting from an initial pose, which attaches the model to one boundary of the tank in the image, the EM algorithm traces these boundaries and converges to a local maximum. The number of iteration steps required to achieve the local maximum changes considerably depending on the provided initial pose.

The resulting pose estimates with respect to the null pose together with the number of iteration steps required for the EM algorithm to converge are given in Table 6.2.

6.3.2 Classification

Pose estimation is crucial in obtaining a predicted image of each model in the data library for best comparison with the actual image. The last stage of the recognition module involves performing these comparisons with each model to get the best alignment which leads to our decision as to the identity of the target in the image.

This comparison is done using the PMPE objective function, which is basically a scaled logarithm of the likelihood function of the pose, defined as the likelihood of our obtaining the observed image features given the true pose vector, β. The objective function is essentially a measure of how closely the projected model features match the image features. Hence, it is expected that the value of this scaled log-likelihood function is greater for the alignment with the correct model.

The values of the objective function evaluated using the final EM alignments of the

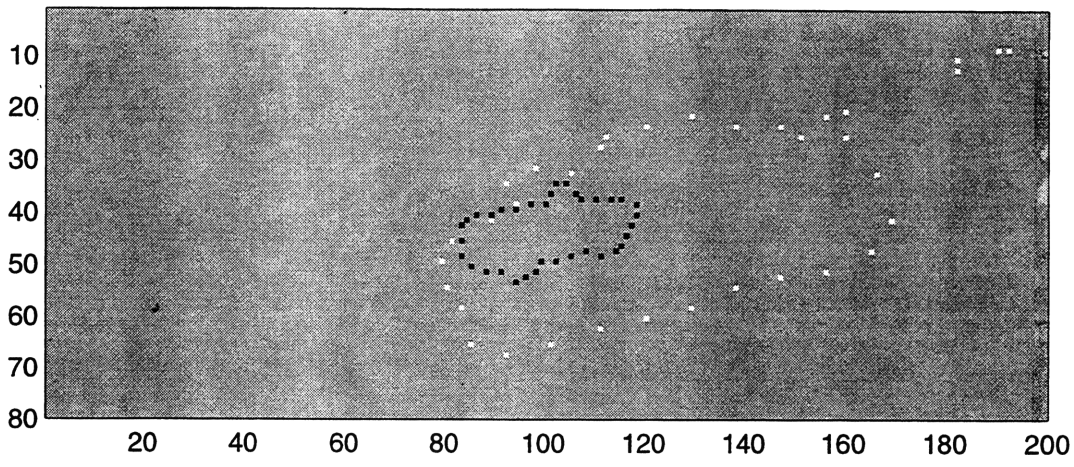

Figure 6-12: Initial alignment with Ford GPA Jeep model.

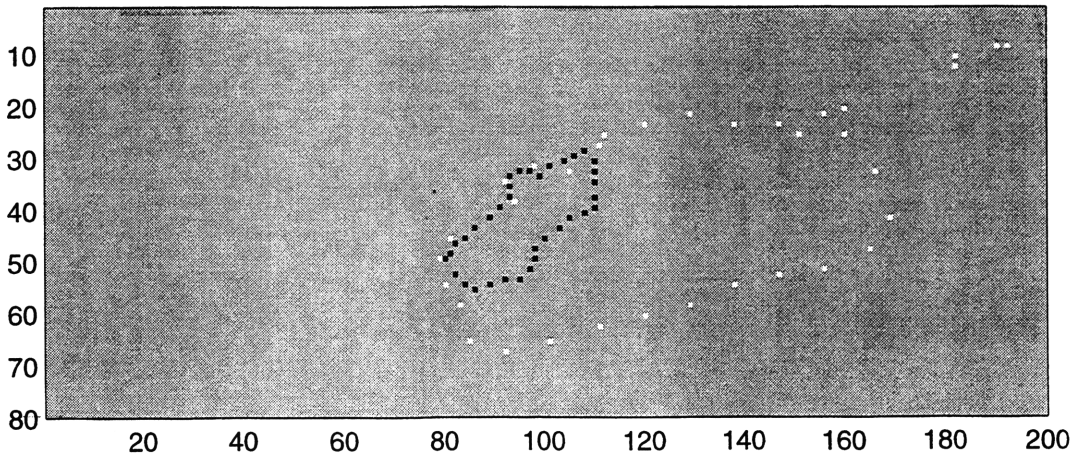

Figure 6-13: Final alignment with Ford GPA Jeep model.

	Null Scores	Initial Scores	EM Scores
M60 Tank Model	63.8718	148.4487	154.5272
T80 Tank Model	56.0621	132.6367	133.9441
GMC CCKW Truck Model	37.4015	76.3891	138.1546
GPA Jeep Model	37.3844	37.3844	60.6572

Table 6.3: Scores for each of the models corresponding to the null pose, hand aligned initial pose and the final EM pose.

object models act as the scores corresponding to each experiment. These scores are compared to deduce the identity of the target among possible candidates known to the recognition system. The final EM scores for our sample image and the four candidate models are given in Table 6.3. Scores corresponding to the null pose and the hand aligned initial pose are also included to show the amount of improvement achieved by the EM refinement step. Comparing these scores leads to our classification of the target in the input image as an M60 tank.

Despite the poor resolution in the input image, the system acquired enough information to arrive at a correct decision for classification. These results imply that mismatch, if it occurs, is mostly likely to be between targets of approximately the same size. The use of actual dimensions in the alignment process thus greatly reduces the probability of misclassification.

6.4 Analyzing the Results

In this section, the alignment results of Figs 6.6 to 6.13 are analyzed. First, multiple trials of matching of the correct model and one of the incorrect models to the image are performed. Then, we investigate the effect of resolution factor in input imagery on the alignment results.

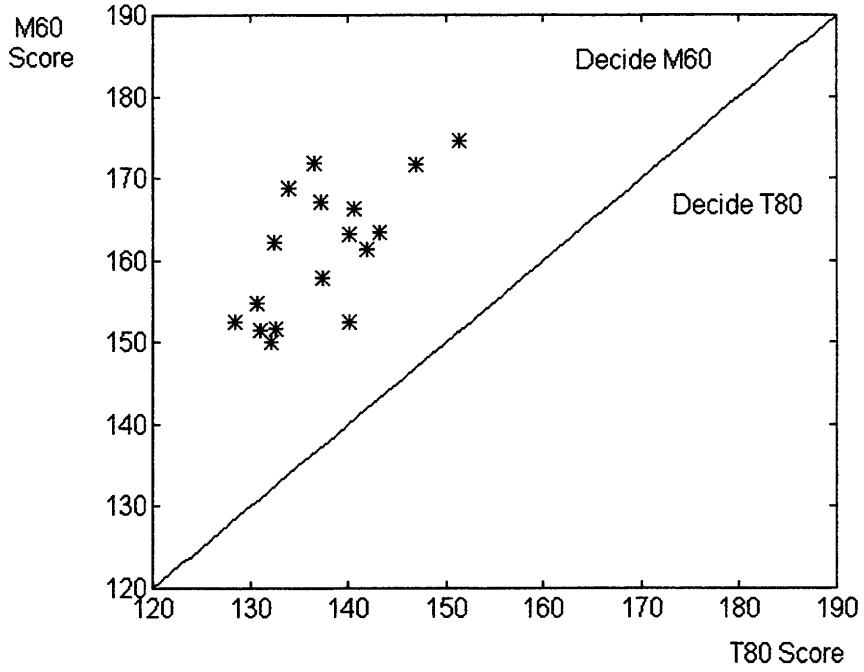

Figure 6-14: Scatter diagram illustrating the scores resulting from multiple-trial experiments.

6.4.1 Multiple Trials

The single-trial experiment conducted in the previous section was able to arrive at the correct decision about the identity of the target in the image. However, this single trial is not adequate to demonstrate the robustness of the algorithm. To be more confident about the behavior of the scores, multiple -trial matching experiments have been conducted for the correct model, the M60 tank model, and one of the incorrect models, the T80 tank model, using simulated data randomly generated from the range truth. Only a small number of trials could be performed due to large run-time requirement of this experiment. The resulting scores obtained from sets of statistically independent identically distributed trials for both of the models are illustrated in Fig. 6-14 in the form of a scatter diagram. The results show that the algorithm gives a correct decision in every case giving us a reasonable level of confidence that the algorithm works properly.

	Mean	Standard Deviation
M60 Tank Model	161.28	7.79
T80 Tank Model	137.49	6.03

Table 6.4: The mean and standard deviation of the scores corresponding to two models

The same set of trials were used to get some understanding of the statistical behavior of the objective function. The statistics may not be quite accurate because only a small number of trials were performed. The mean and the standard deviation for the two distinct models estimated using these trials are shown in Table 6.4. Figs. 6.15 and 6.16 are illustrations of the 17-trial histograms for the objective function values for the correct model and the incorrect model.

6.4.2 Resolution Factor

An additional analysis was done on the recognition experiments to deduce the effect of spatial resolution on the alignment and scoring process.

The input preprocessor, namely the fast EM/ML algorithm, can profile noisy range imagery involving anomalous pixels at arbitrary levels of spatial resolution. To test the effect of spatial resolution of input data on the alignment results, a noisy raw range data was generated from the range truth and profiled at 2×2-pixel and 4×4pixel block sizes. Moreover, we have also used the range truth as the input to our system, assuming that it is the profiled image corresponding to a higher resolution noisy raw range data. Using these three preprocessed images we can probe the effect of increased resolution on the performance of our alignment and scoring algorithm.

Each of these three restored range images are processed to extract compact information required for matching, i.e., all are transformed into feature domain where the matching with the model features of the correct model is performed. The corresponding alignment results are seen in Fig. 6.17, where the top figure corresponds to the alignment with the range truth, the middle figure with the 2×2 profile and the bottom figure with the 4×4 profile.

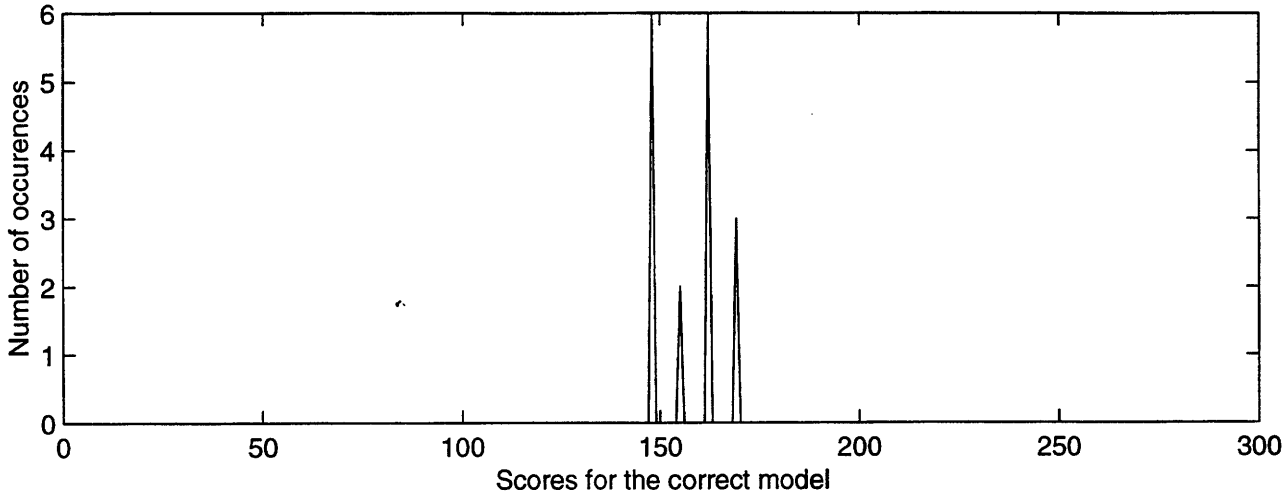

Figure 6-15: Histogram of scores resulting from alignments with M60 A3 tank model.

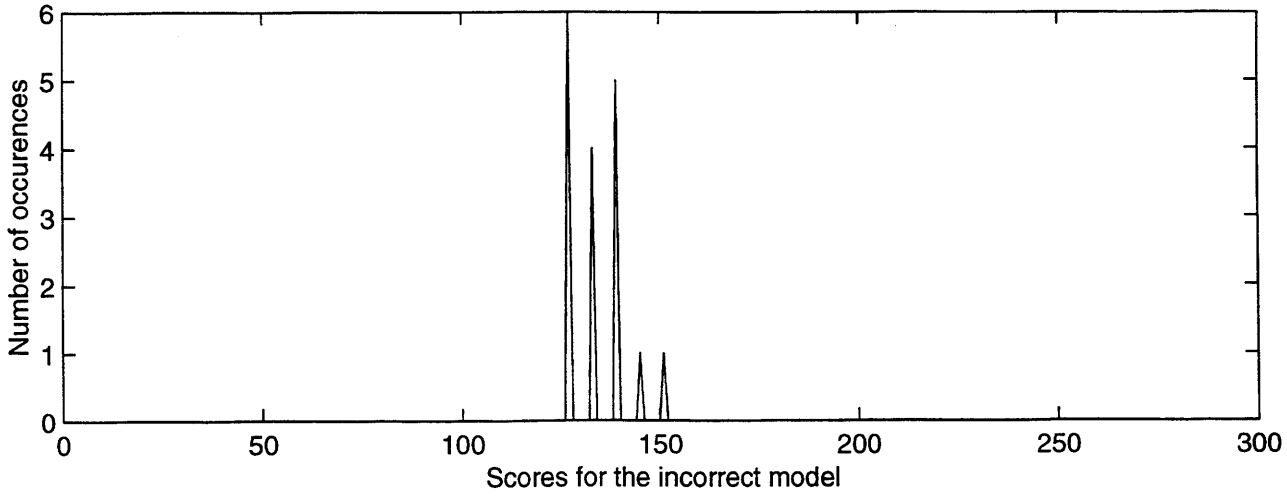

Figure 6-16: Histogram of scores resulting from alignments with T80 tank model.

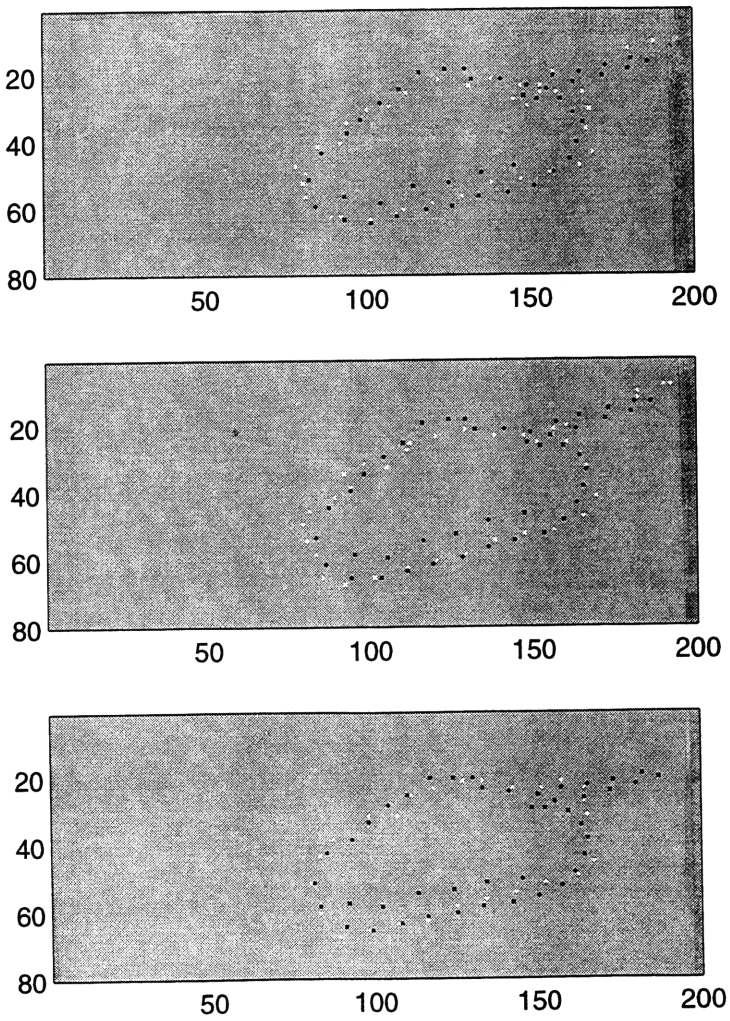

Figure 6-17: Alignment results of the correct model with different resolution input data: top 1×1 block size, middle 2×2 block size, bottom 4×4 block size.

	Final Scores	Number of Features	Number of Iteration Steps
Range Truth	164.5077	38	6
2×2 Profile	154.5272	37	10
4×4 Profile	130.3048	29	9

Table 6.5: Scores and relevant data corresponding to matching input data of different resolutions with the correct model.

The resulting scores of alignment together with the related data used in the matching process are given in Table 6.5. The results clearly indicate that the resolution associated with the input data has a strong impact on the scores, which reflect the degree of match between the two sets of features. This is expected since more detailed information extracted from the input data about the target of interest will result in a better match with the object model. Note that the number of features extracted decreases as the resolution is reduced in the restored images consistent with the reduction in the level of detail.

A similar experiment has been performed to test the effect of spatial resolution of input imagery on the alignment results of the target with an incorrect model, the T80 Tank Model. In particular, three images preprocessed at different levels of resolution are matched with the T80 Tank Model and the corresponding scores are compared. In the previous section, it has been demonstrated that our algorithm gives a correct decision between the M60 Tank model and the T80 Tank Model when the noisy input image is profiled using 2×2-pixel blocks. By means of this experiment, we can probe the effect of spatial resolution of input images on the classification ability of our algorithm.

The resulting alignment results are illustrated in Fig. 6.18 and the scores of alignment of these three input images with the T80 Tank model are given in Table 6.6 with the relevant data. The results show that decreasing the spatial resolution decreased the alignment scores for the incorrect model as well. This behavior can be explained by examining the alignment results. As the block size in profiling the range data is increased, the target edges tend to be characterized by sharper corners and straight lines. This results in a poorer matching with the incorrect tank model, because it has smooth curved edges.

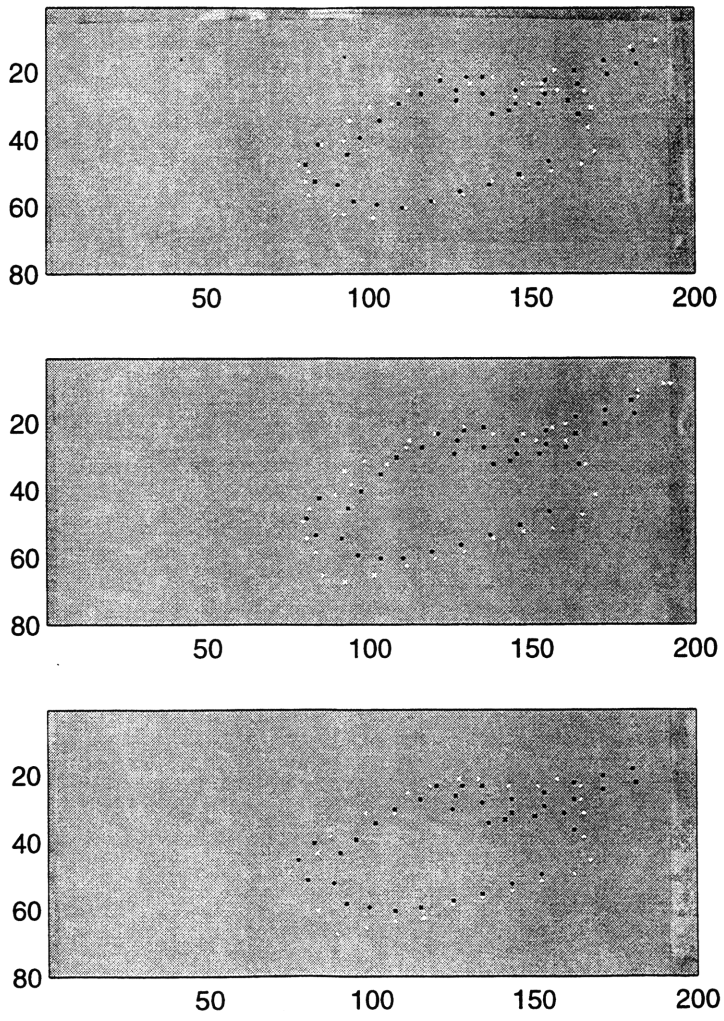

Figure 6-18: Alignment results of the incorrect model with different resolution input data: top 1×1 block size, middle 2×2 block size, bottom 4×4 block size.

	Final Scores	Number of Features	Number of Iteration Steps
Range Truth	147.684	38	6
2 × 2 Profile	133.9441	37	6
4 × 4 Profile	103.7614	29	13

Table 6.6: Scores and relevant data corresponding to matching input data of different resolutions with the incorrect model.

	Final Scores for the Correct Model	Final Scores for the Incorrect Model
Range Truth	164.5077	147.684
2×2 Profile	154.5272	133.9441
4×4 Profile	130.3048	103.7614

Table 6.7: Scores for matching input data of different resolutions with the correct and the incorrect model.

Table 6.7 displays the final alignment scores for the three preprocessed images for the correct and the incorrect model together. As seen, despite change in resolution of input imagery, the algorithm is still successful in reaching at correct decisions about the identity of the target.

6.4.3 Feature extraction mechanism

Our next goal is to assess the effects of sensor physics capabilities on the feature extraction mechanism and hence on object recognition system performance. This is a complicated problem since the statistical behavior of the features extracted from the target in the image is affected by a variety of mechanisms. The variance for feature deviations perpendicular to the underlying edge curve is a good measure of object recognition performance because the matching algorithm essentially examines the deviations of the actual and the predicted features and the resulting alignment score is a function of these deviations. In this section, we first derive an analytical expression for this variance under some reasonable simplifying assumptions. We then generate synthetic range images and calculate the variance experimentally for different values of sensor parameters.

Analytical Results

The matching algorithm in our system uses edge-based features to make the comparison between the actual and predicted images. Thus the variability in the feature location is equivalent to the variability in the edge detection.

It is crucial that we identify the mechanisms that result in uncertainties in the edge

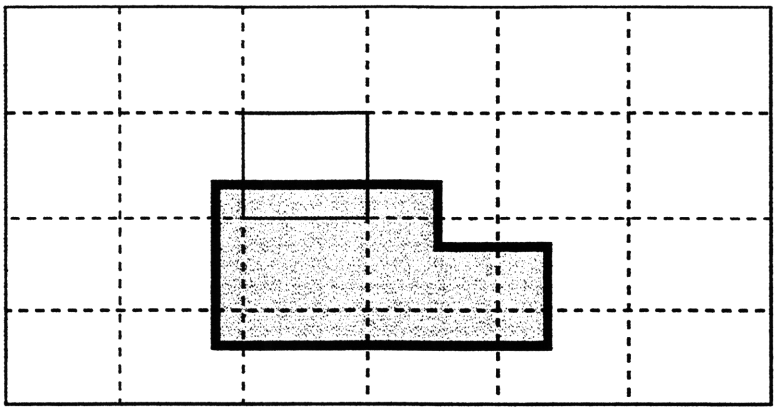

Figure 6-19: Effect of beam width on edge detection process.

detection process. One of the effects is the beam width of the laser radar, which determines the range measurement per pixel. At the edges of the target, the field-of-view imaged by the beam may include both the target and the background behind it, as shown in Fig. 6.19. The range measurement for that pixel is determined based on the relative intensities of the return signals which are proportional to the relative areas of the target and the background regions viewed by the beam. Other effects that may cause variability in detecting the edges in the range image are the speckle behavior, which results in range anomalies, and the local oscillator shot noise, which results in Gaussian noise in the local accuracy of the range measurements.

It is possible to develop a probability density function, which incorporates these effects, for the variation of a point feature on the edge perpendicular to the underlying edge curve. The probability density function takes a complicated form for edges which are not aligned with the pixel grid. However for pixels on aligned edge contours, it simply takes the form of a uniform distribution since given the pixel grid, the edge is equally likely to be anywhere in the image and the area covered by the target in the region viewed by the beam, which determines the range measurement for that pixel, is a linear function of

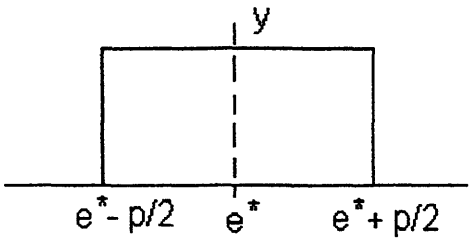

Figure 6-20: Perpendicular feature deviation is characterized by a uniform distribution centered on the correct location, e^*, with support equal to the pixel size, p.

the perpendicular variation of the edge. This distribution is centered at the correct edge point, denoted by e^*, and has a support equal to the pixel size, as illustrated in Fig 6-20, where y is a random variable representing the detected location of the edge point and p is a random variable representing the pixel size. Note that the pixel size is a random variable since it is a function of the measured range value. The variance of y is given by

$$\begin{aligned} Var(y) &= E\left[Var(y|p)\right] + Var\left(E[y|p]\right) \\ &= E\left[\frac{p^2}{12}\right] + \underbrace{Var(e^*)}_{0} \\ &= E\left[\frac{p^2}{12}\right] \end{aligned} \tag{6.11}$$

The pixel size p is a function of the measured range value

$$p = \phi r \tag{6.12}$$

where r is the measured range value in meters and ϕ is the full-angle beam width of the laser radar. The expected value and variance of p can be found in terms of the expected value and the variance of r,

$$
\begin{aligned}
Var(y) &= E\left[\frac{p^2}{12}\right] \\
&= \frac{1}{12}\left[Var(p) + \left(E[p]^2\right)\right] \\
&= \frac{1}{12}\left[\left(\phi^2 Var(r)\right) + \left(\phi E[r]\right)^2\right] \qquad (6.13)
\end{aligned}
$$

It is shown in previous work [1] that the optimal front end processor used in our system, the fast EM/ML algorithm, calculates range estimates that are nearly unbiased with error variances that approach the complete-data bound at higher resolutions, which is slightly weaker than the Cramér-Rao bound, the ultimate performance limit of unknown parameter estimation. Therefore it is reasonable to assume

$$E[r] = r^* \qquad (6.14)$$

$$Var(r) = \frac{\delta R^2}{1 - \Pr(A)} \frac{P}{Q} \qquad (6.15)$$

where P is the resolution, Q is the number of pixels in the range image, δR is the local range accuracy and $Pr(A)$ is the probability of anomaly. δR and $\Pr(A)$ represent the physical parameters of a real laser radar system. Substituting these values in Eq. 6.13 leads to

$$
\begin{aligned}
Var(y) &= \frac{1}{12}\left[\phi^2\left(\frac{\delta R^2}{1 - \Pr(A)}\frac{P}{Q}\right)\right] + \frac{1}{12}\left(\phi r^*\right)^2 \\
&= \frac{1}{12}\left[\phi^2\left(\frac{\delta R^2}{1 - \Pr(A)}\frac{P}{Q}\right)\right] + \frac{1}{12}(p^*)^2 \qquad (6.16)
\end{aligned}
$$

where p^* is the pixel size at the true range value of the target.

Table 6.8 shows the resulting feature variance and standard deviation for certain

Coarsening	Pixel Size	$Pr(A)$	δR	Variance	Standard Deviation
8×8	0.16 m	0.05	2 bins	21.3 cm^2	4.61 cm
10×10	0.2 m	0.05	2 bins	33.3 cm^2	5.8 cm
8×8	0.16 m	0.1	4 bins	$\sim$21.3 cm^2	$\sim$ 4.62 cm

Table 6.8: Variance and standard deviation for feature fluctuation for different values of sensor parameters.

values of sensor parameters. It is clear from these results that the pixel size, which is a function of the laser beam width and the range of the target to the sensor, is the dominant factor which affects the feature fluctuation and in turn the object recognition performance. The anomalies and the Gaussian noise that affects the range estimates are not that significant, because of the full angle beam width which is in general on the order of 0.1 mrad.

Experimental results

We would like to study the performance of the object recognition system we developed as a function of sensor parameters using feature variance as a measure. The real input imagery used in this work has been taken from Infrared Airborne Radar (IRAR) data release. Unfortunately, the available real images are inadequate to perform experiments in which the sensor parameters can be varied arbitrarily. Therefore, we have generated synthetic range images using 3-D CAD models of the targets.

We are particularly interested in the statistical behavior of features along edges that align with the pixel grid so that we can use these experimental results to confirm the validity of the statistical model developed in the previous section. Therefore, we used the GMC CCKW truck model which has a fairly flat top to generate 2-D rendered views in which the top aligns with horizontal pixel grid. The result is a 464×454 binary image as seen in Fig. 6-21. We then generated a planar background whose slope is chosen according to the slope of the planar background associated with the real laser radar range image used in the recognition experiments. The target is placed at 500 meters on this plane. The range values in the generated image vary between 450 to 580 meters

approximately. These values are converted to range bins by subtracting the range gate offset of 427 meters and quantizing the values to 1.1 meters to get an equivalent image in terms of range bins, as shown in Fig. 6-22. This corresponds to coarsening the range resolution and has the effect of losing information about the ground attachment line because of poor range contrast. This image is between 0 and 150 range bins as indicated in the calibration bar on the right.

The spatial resolution of the resulting image is coarsened by choosing 8×8-pixel blocks and assigning to each block a value consistent with the range values of the majority of the pixels inside the block. Finally, we add anomalous pixels at 5% rate to simulate the speckle behavior and add zero-mean Gaussian noise with standard deviation, $\delta R = 2$, to each pixel to account for the local oscillator shot noise similar to what we have done for the real laser radar range imagery. We have set the range values of the pixels at the edges to zero to simulate the effect of dropouts that appear on the edges of real imagery. The final synthetic range image is illustrated in Fig. 6-23.

The generated synthetic range image is applied as the input to our system and matched with the original GMC CCKW Truck Model. The resulting alignment of image features with the model features is shown in Fig. 6-24. In this figure the white dots represent the image features whereas the black dots represent the model features and the pixel size is 0.1 meters. This procedure is performed ten times using simulated data randomly generated from the coarsened synthetic range image by adding noise and anomalous pixels. The variance for the perpendicular feature deviation is calculated as the sample variance of the features at the top of the truck.

This experiment is repeated by changing the sensor parameters: first by increasing the standard deviation of the added Gaussian noise to 4 range bins and $Pr(A)$ to 0.1 and then by coarsening the image by 10×10-pixel blocks. The calculated variances and standard deviations for these cases are shown in Table 6.9 together with the sensor parameters used in the experiments.

These results confirm our earlier observation that the feature variance is affected

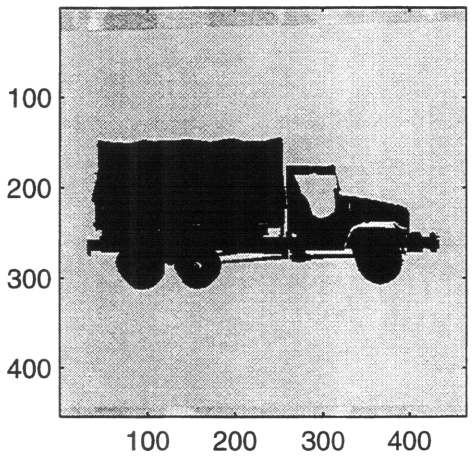

Figure 6-21: Rendered view of GMC CCKW truck model.

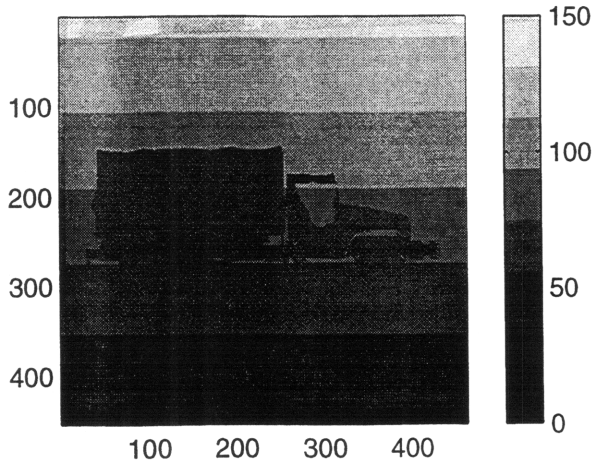

Figure 6-22: Noiseless synthetic range image of a truck with a planar background, coarsened in range resolution.

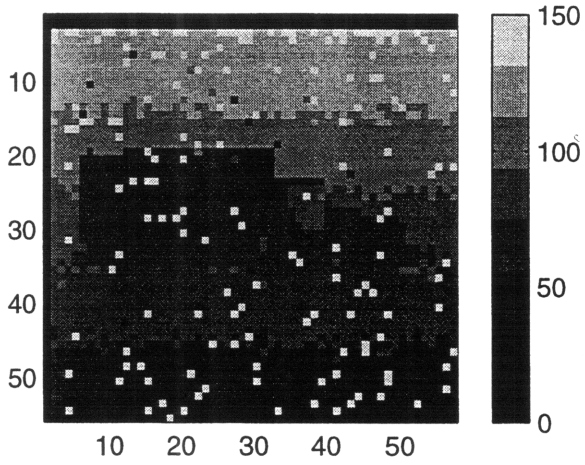

Figure 6-23: Synthetic range image of a truck.

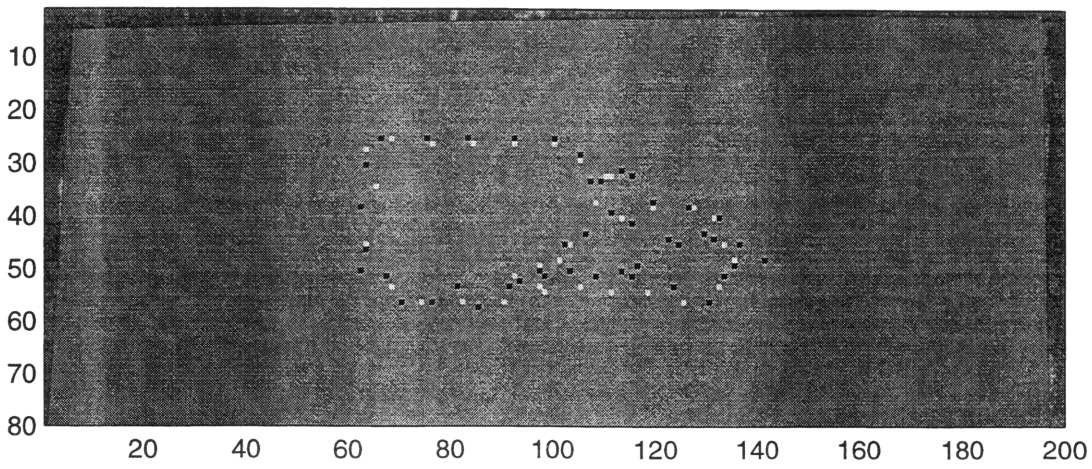

Figure 6-24: Alignment of the image and the model features.

Coarsening	Pixel Size	$Pr(A)$	δR	Variance	Standard Deviation
8×8	0.16 m	0.05	2 bins	11.1 cm^2	3.33 cm
10×10	0.2 m	0.05	2 bins	305.2 cm^2	17.47 cm
8×8	0.16 m	0.1	4 bins	12.48 cm^2	3.53 cm

Table 6.9: Variance and standard deviation for feature fluctuation for different values of sensor parameters

mainly by the pixel size, which in turn is a function of beam width and range of the target to the sensor. The effect of the noise and anomalous pixels is minor. The experimentally calculated values differ from the analytically calculated values, which may be due to two reasons. First, the variance of the range estimate approaching the complete-data bound is an approximation used in the analytical calculation. The range estimate variance may be affected by the resolution of the input imagery. Secondly, the number of trials performed could have been insufficient to obtain an accurate experimental estimate of the variance.

Chapter 7

Conclusions

Recognizing 3-D objects from range imagery has received considerable attention in the last few years [21,22]. Laser radar range imagery taken from IRAR data set is characterized by coarse range precision and low azimuth and elevation resolution. In addition, it is degraded by the combined effects of laser speckle and local oscillator shot noise, resulting in range anomalies and Gaussian noise in the local accuracy of the range measurements. In this thesis, our objective was to develop a statistically optimum approach for doing model-based object recognition using low-resolution, degraded laser radar range images. We have attempted to build an end-to-end system that operates in an autonomous fashion, using raw sensor images as its input and making a recognition decision at the output.

In the computer vision literature, many of the approaches used in object recognition focus on target detection and pose estimation [5,24]. However, our goal was to do object recognition in the general sense, i.e., our algorithm was supposed to deduce the identity of the target in the image among possible candidates in a database. Therefore, the first step in constructing the system was forming a data library involving 3-D CAD models which are adequate representations of military targets that are likely to be present on the site imaged by the laser radar. The system is built on the basic idea of providing a score for the degree of match between the image of interest and each of the models and comparing these results to reach a final decision.

In developing the system, we used a modular and well-understood building block approach. Our intent was to use established and mathematically well-understood techniques, especially in the main components of the system. However, we did use ad hoc techniques, as in feature extraction process, when necessary. The structure of the developed object recognition system can be decomposed into preprocessing, segmentation, feature extraction and matching steps. Throughout our development, we investigated the problem in two separate stages: processing the input data to extract compact information and matching.

The preprocessor is the first component in processing the raw input data. Its aim is to reduce the sensor-dependent perturbations so that the image quality is enhanced, increasing the performance of the successive modules. For the preprocessor, we have used the fast EM/ML algorithm. It performs maximum-likelihood range profile estimation via the expectation-maximization algorithm, subject to regularity conditions. This approach differs from ad hoc image enhancement techniques by the virtue of building on the sensor physics and thus providing the required, quantified near-optimal performance characteristics. By using the special structure of Haar wavelet basis in the maximization step of the EM algorithm, a computationally efficient and numerically robust procedure is obtained. The next step in our system is segmentation. Segmentation is employed not because of clutter in the image, but to deal with the ground attachment of the target. There are many different approaches that can be employed to this segmentation task. However, because the input imagery is a planar, sloping, featureless background with a target embedded in it, we have found employing an optimal laser radar imaging approach to be the most appropriate to satisfy our purpose. Basically, we have estimated the target and background planes separately, using the ML/EM planar range profiling approach, and isolated the target region from the background as accurately as possible. The last step prior to pose estimation is to extract features of both the object in the restored image and the object model. In both cases this is accomplished from the edge curves determining the boundaries. What results is two sets of final, compact information for

use in the matching process.

The matching step uses a statistical approach, posterior marginal pose estimation (PMPE), to achieve and score the best alignment between the image and the model data. The resulting objective function is optimized using the EM algorithm to find the best pose estimate starting from a good initial pose, i.e., we assumed that we are provided with the initial pose value and our purpose is to refine this value using the EM algorithm.

The search process is simplified by splitting the six-dimensional pose space into two parts. The first part covers the two out-of-plane rotation pose parameters, which we determine by forming a set of discrete hypotheses in the form of 2-D synthetic views of the 3-D model. The use of available range information then reduces the dimension of the search space to three, involving in-plane translations and in-plane rotation. In our work, we focus on these three parameters. Reduction of the pose space, in addition to being computationally efficient, brings an additional benefit. Both the image features and the model features are transformed into a new coordinate system, using the available data, in actual object dimensions instead of number of pixels. As a result, models of considerably different sizes with respect to the actual target in the image are automatically eliminated because they are not allowed to be scaled, and hence are constrained to have poor degrees of alignment. The only misclassification possibilities arise from models having approximately the same size as the target.

The recognition experiments performed to test the system confirm these expectations. These experiments show that, despite the low-resolution nature of the input imagery, our algorithm succeeds in gaining adequate information to correctly recognize the target among the military vehicles known to the recognition system. It is also possible to discriminate two different models of the same vehicle, i.e., two tanks, by means of this algorithm, as long as their shapes and sizes are sufficiently distinct.

Recognition performance was explored by employing multiple trials of matching of the correct model and one of the incorrect models to randomly generated range data. These results also achieved the correct decision in each case, giving us confidence in the

robustness of the developed system. Further analysis involved performing matching of the correct and the incorrect tank models to input imagery preprocessed at different levels of resolution. The results demonstrated that the alignment performance was degraded by reducing the spatial resolution of the input imagery. The target edges become sharper by increasing the block size used in preprocessing, resulting in poorer matches with the tank models having curved edges.

Finally, the object recognition performance was investigated in terms of the effect of sensor parameters. An analysis, using reasonable assumptions, followed by experiments in which synthetic range images were generated, was used to explore the effect of varying the sensor parameters. The results show that the most important parameters are the laser beam width and the range of the target to the sensor, which collectively determine the size of the pixels on the target. The effect of the anomalous pixels and the Gaussian noise is not that significant, because of the size of the beam width and our use of the optimal front-end processor whose range-estimation performance approaches the complete-data bound.

Future work can proceed in many directions. First, it is necessary to find a method to figure out a suitable initial pose for PMPE refinement. In principle this could be achieved by an exhaustive search over the whole image. Determining the initial pose becomes even more difficult when clutter is involved in the image. Our algorithm claims that the misclassification occurs only because of the object models in the data library that are similar in appearance and size to the correct model associated with the input data. The algorithm can be improved by including a hypothesis testing step after the classification is done to reject the targets having scores below a certain threshold. Recognition may also be improved by identifying appendage-like parts which are highly specific to particular vehicles. In addition to these issues, there is room for improvement in the segmentation and feature extraction steps of the system so that they provide better performance.

Bibliography

[1] T.J. Green, Jr. and J.H. Shapiro, "Maximum-likelihood laser radar range profiling with the expectation-maximization algorithm," *Opt. Eng.* **31**, 2343-2354 (1992).

[2] D.R. Greer, "Multiresolution laser radar range profiling of real imagery," M. Eng. Thesis, Dept. of Elect. Eng. and Comput. Sci., MIT, January 1996.

[3] I. Fung, "Multiresolution laser radar range profiling with the EM algorithm," S.M Thesis, Dept. of Elect. Eng. and Comput. Sci., MIT, May 1994.

[4] D.R. Greer, I. Fung, and J.H. Shapiro, "Maximum-likelihood multiresolution laser radar range imaging," *IEEE Transactions on Image Processing*, Vol. 6, No. 1, 36-46 (Jan. 1997).

[5] W.M. Wells III, "Statistical object recognition," Ph.D. Thesis, Dept. of Elect. Eng. and Comput. Sci., MIT, February 1993.

[6] J.H. Shapiro, R.W. Reinhold, and D. Park, "Performance analyses for peak detecting laser radars," Proc. SPIE **663,** 38-56(1986).

[7] M.B. Mark and J.H. Shapiro, "Multipixel, multidimensional laser radar system performance," Proc. SPIE 783, 109-122 (1987).

[8] S.M Hannon and J.H. Shapiro,"Laser radar target detection with a multipixel joint range-intensity processor," Proc. SPIE 999, 162-175 (1988).

[9] S.M Hannon and J.H. Shapiro, "Active-passive detection of multipixel targets," Proc. SPIE 1222, 2-23 (1990).

[10] T.J. Green, Jr. and J.H. Shapiro, "Detecting objects in three-dimensional laser radar range images," *Opt. Eng.* **33,** 865-874, (1994).

[11] T.J. Green, Jr., J.H. Shapiro, and M.M. Menon, "Target detection performance using 3-D laser radar images," Proc. SPIE, 328-341 (1991).

[12] R.C. Harney and R.J. Hull, "Compact infrared radar technology," Proc. SPIE **227**, 162-170, (1980).

[13] A.B. Gschwendtner, R.C. Harney, and R.J. Hull, "Coherent IR radar technology," in D.K Killinger and A. Mooradian, eds., *Optical and Laser Remote Sensing* (Springer-Verlag, Berlin, 1983).

[14] J.W. Goodman, "Statistical properties of laser speckle," in J.C. Dainty, ed., *Laser Speckle and Related Phenomena* (Springer-Verlag, Berlin, 1975).

[15] R.H. Kingston, *Detection of optical and Infrared Radiation* (Springer-Verlag, Berlin,1978). Chap. 3.

[16] A.P. Dempster, N.M Laird, and D.B Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser.* **39**, 1-38 (1977).

[17] J.D. Baker, and, W.M. Wells III, "Multiresolution statistical object recognition," *Proc. of Image Understanding Workshop,* (November 1994).

[18] W.M. Wells III, "Statistical approcahes to feature-based object recognition," *International Journal of Computer Vision* **21**, 63-98 (1997).

[19] W.M. Wells III, "Statistical object recognition with the expectation-maximization algorithm in range-derived features," *Proc. of the DARPA Image Understanding Workshop*, 839-850 (April 1993).

[20] J.K. Bounds, *The Infrared Airborne Radar Sensor Suite, RLE Technical Report* **610**, (December 1996).

[21] J.G. Verly, R.L. Delanoy, and D.E. Dudgeon, "Model-based system for automatic target recognition from forward-looking laser-radar imagery," *Opt. Eng.* **31(12)**, 2540-2552, (1992).

[22] J.G. Verly, R.L. Delanoy, and D.E. Dudgeon, "Machine intelligence technology for automatic target recognition," *The Lincoln Laboratory Journal* **2(2)**, 277-307, (1989).

[23] C.F. Olson, and D.P. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Transactions on Image Processing*, Vol. 6, No. 1, 103-113, (Jan. 1997).

[24] J. Hornegger, and H. Niemann, "Statistically optimal model estimation and object recognition," unpublished.

[25] S. Ullman, and R. Basri, "Recognition by linear combinations of models," *A.I.Memo* **1152**, Massachusetts Institute of Technology, (1989).